

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 08:20:25 ; Search time 93 Seconds
(without alignments)
2111.511 Million cell updates/sec

Title: US-10-719-885-2
Perfect score: 3669
Sequence: 1 MKRGKDEKLEMPFPLHV.....AARIFQSIQERKRYDSSKP 695

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A Geneseq 29Jan04:*
1: geneseq1980s:*
2: geneseq1990s:*
3: geneseq2000s:*
4: geneseq2001s:*
5: geneseq2002s:*
6: geneseq2003as:*
7: geneseq2003bs:*
8: geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3669	100.0	695	3 AAY45097	Aay45097 Arabidops
2	3327	90.7	695	6 AAO15964	Aao15964 Arabidops
3	3321	90.5	695	6 AAO15963	Aao15963 Arabidops
4	197	5.4	971	7 ADE56304	Ades56304 Human Pro
5	193.5	5.3	802	5 ABG93215	Abg93215 S. cerevi
6	193.5	5.3	802	6 ABR53105	AbR53105 Protein s
7	192	5.2	1503	4 ABB60944	Abb60944 Arabidops
8	189.5	5.2	738	5 ABG93140	Abg93140 S. cerevi
9	189.5	5.2	1130	5 ABR73921	AbR73921 Candida a
10	186.5	5.1	1138	5 ABR705434	AbR705434 Human bre
11	186	5.1	771	4 ABB64977	Abb64977 Arabidops
12	180	4.9	1594	4 ABB63688	Abb63688 Arabidops
13	179.5	4.9	1133	7 AAE36097	Aae36097 Human Mas
14	179.5	4.9	1133	7 ADC59316	Adc59316 Human pol
15	178	4.9	1420	4 ABB63410	Abb63410 Arabidops
16	177	4.8	748	5 ABB92874	Abb92874 Arabidops
17	176	4.8	2237	4 ABB64977	Abb64977 Arabidops
18	174.5	4.8	757	4 AAG85017	Aag85017 Shrimp wh
19	172	4.7	1187	4 ABB67666	Abb67666 Arabidops
20	172	4.7	3502	4 ABB58382	Abb58382 Arabidops
21	171.5	4.7	1596	6 AAE36095	Aae36095 Arabidops
22	170.5	4.6	843	4 AAM79967	Aam79967 Human pro
23	170	4.6	2328	4 ABB71657	Abb71657 Arabidops
24	168.5	4.6	333	4 ABB62536	Abb62536 Arabidops
25	168.5	4.6	506	4 ABB71899	Abb71899 Arabidops

26	168.5	4.6	1125	7 ADC37561	Adc37561 Human nuc
27	166.5	4.5	1441	7 ADC23487	Adc23487 Receptor
28	165.5	4.5	1366	4 ABB66921	Abb66921 Arabidops
29	165.5	4.5	1920	4 ABB64441	Abb64441 Arabidops
30	165	4.5	611	5 ABG93274	Abg93274 C. albica
31	165	4.5	1790	4 ABB71739	Abb71739 Arabidops
32	164.5	4.5	520	3 AAG36717	Aag36717 Arabidops
33	164.5	4.5	571	3 AAG36716	Aag36716 Arabidops
34	164.5	4.5	629	3 AAG36715	Aag36715 Arabidops
35	164.5	4.5	2703	4 ABB63299	Abb63299 Arabidops
36	164.5	4.5	2703	5 AAG70019	Abb70019 Larval vi
37	163.5	4.5	522	3 AAG13630	Aag13630 Arabidops
38	163.5	4.5	539	3 AAG13629	Aag13629 Arabidops
39	163.5	4.5	540	3 AAG13628	Aag13628 Arabidops
40	163.5	4.5	821	4 AAM78983	Aam78983 Human pro
41	163.5	4.5	2237	5 AAG70004	Abb70004 Larval vi
42	163.5	4.5	2309	4 ABB66232	Abb66232 Arabidops
43	163.5	4.5	2703	4 ABB60074	Abb60074 Arabidops
44	163	4.4	688	4 ABB67404	Abb67404 Arabidops
45	163	4.4	690	4 ABB66052	Abb66052 Arabidops
46	163	4.4	1128	4 ABB58911	Abb58911 Arabidops
47	163	4.4	3080	4 ABB64877	Abb64877 Arabidops
48	162.5	4.4	2017	6 ABB54584	Abb54584 Human NOV
49	162	4.4	749	4 ABB62915	Abb62915 Arabidops
50	162	4.4	816	2 AAR71111	Aar71111 Spinocere
51	162	4.4	816	2 AAY33494	Aay33494 Human SCA
52	162	4.4	816	6 ABR43117	AbR43117 Human ata
53	161.5	4.4	2062	7 ADC31187	Adc31187 Human nov
54	161	4.4	971	2 AAW48896	Aaw48896 Candida a
55	161	4.4	1322	4 ABB60275	Abb60275 Arabidops
56	161	4.4	1537	4 ABB66989	Abb66989 Arabidops
57	160.5	4.4	1441	5 AAE12570	Aae12570 Human ste
58	160.5	4.4	1441	5 AAE25569	Aae25569 Human ste
59	160.5	4.4	1441	6 AAE36003	Aae36003 Human SRC
60	159	4.3	993	4 ABB64403	Abb64403 Arabidops
61	159	4.3	3201	4 ABB62899	Abb62899 Arabidops
62	158.5	4.3	459	5 ABR73940	AbR73940 Candida a
63	158.5	4.3	926	4 ABB65135	Abb65135 Arabidops
64	158.5	4.3	1061	2 AAW26370	Aaw26370 Human ste
65	158.5	4.3	1494	5 AAU78460	Aau78460 Mouse bet
66	158	4.3	3572	5 AAG95659	Abb95659 Human nuc
67	157.5	4.3	296	7 ADD45442	Add45442 Rat Prote
68	157.5	4.3	296	7 ADE57105	AdE57105 Rat Prote
69	157.5	4.3	1566	6 ABR59716	AbR59716 Human RER
70	157	4.3	1424	4 AAM39253	Aam39253 Human pol
71	157	4.3	1586	6 ABU21467	Abu21467 Protein e
72	156.5	4.3	594	4 ABB61362	Abb61362 Arabidops
73	156.5	4.3	700	4 ABB61671	Abb61671 Arabidops
74	156.5	4.3	3238	4 ABB71715	Abb71715 Arabidops
75	156	4.3	531	3 AAB56866	Aab56866 Human pro
76	156	4.3	1703	4 ABB66223	Abb66223 Arabidops
77	155.5	4.2	1367	7 ADB70311	AbD70311 C. neofo
78	155	4.2	392	4 AAG10550	Abb10550 Novel hum
79	155	4.2	404	2 AAR90675	Aar90675 Human chr
80	155	4.2	647	4 ABB67093	Abb67093 Arabidops
81	155	4.2	672	4 ABB59473	Abb59473 Arabidops
82	155	4.2	1153	6 AAG79910	Aag79910 MAML2. 4/
83	155	4.2	1153	7 ADC59310	Adc59310 Human pol
84	155	4.2	1445	4 ABB64619	Abb64619 Arabidops
85	154.5	4.2	1427	4 ABB60419	Abb60419 Arabidops
86	154.5	4.2	1447	2 AAW81029	Aaw81029 Murine pc
87	154.5	4.2	1902	6 ABB38695	Abb38695 Human nuc
88	154	4.2	530	3 AAG40002	Aag40002 Arabidops
89	154	4.2	581	3 AAG40001	Aag40001 Arabidops
90	154	4.2	639	3 AAG40000	Aag40000 Arabidops
91	153.5	4.2	502	3 AAG24542	Aag24542 Arabidops
92	153.5	4.2	1381	5 AEG93234	Aeg93234 C. albica
93	153.5	4.2	1381	5 ABR73947	Abb73947 Candida a
94	153	4.2	274	7 ADD47240	Add47240 Rat Prote
95	153	4.2	274	7 ADD48703	Add48703 Human pol
96	153	4.2	1464	4 AAM41039	Aam41039 Human pol
97	152.5	4.2	1101	6 ABR47437	AbR47437 Breast ca
98	152	4.1	349	4 ABB60321	Abb60321 Arabidops

99	152	4.1	724	4	ABB65778	Abb65778 Drosophil
100	152	4.1	1012	2	AAY17406	Aay17406 Human atr
ALIGNMENTS						
RESULT 1						
ID	AA45097	standard; protein; 695 AA.				
XX	AA45097;					
DT	05-JUN-2000	(first entry)				
DE	Arabidopsis thaliana early-flowering protein, ELF3.					
XX	ELF3; early-flowering; circadian clock; photoperiodism; elf3 mutant;					
KW	photoperiod-insensitive; hypocotyl elongation; transgenic plant; corn;					
KW	rice; barley; cotton.					
XX	Arabidopsis thaliana.					
OS	WO200009658-A2.					
XX	24-FEB-2000.					
XX	17-AUG-1999;	99WO-US018747.				
XX	17-AUG-1998;	98US-0096802P.				
XX	(UYOR-) UNIV OREGON.					
PI	Wagner R, Hicks KA, Spence MZ, Albertson TM, Foss H, Prigge M;					
XX	WPI; 2000-224320/19.					
DR	N-PSDB; AA251052, AA251053.					
XX	Novel ELF3 gene encoding plant proteins involved in photoperiodism and					
PT	circadian rhythms used to alter photoperiodic and/or circadian clock-					
PT	based gene expression of plants.					
XX	Claim 1a; Page 34-36; 45pp; English.					
PS	The present amino acid sequence is the Arabidopsis ELF3 (early-flowering)					
XX	protein, that regulates the circadian clock functions and photoperiodism					
CC	in plants. The elf3 mutants are photoperiod-insensitive and have long-					
CC	hypocotyl, since they are defective in blue-light dependent inhibition of					
CC	hypocotyl elongation. The ELF3 protein can complement an elf3 mutant. The					
CC	ELF3 sequences are used to alter photoperiodic and circadian clock- based					
CC	gene expression in transgenic plants. Such transgenic plants include					
CC	Arabidopsis, pepper, tomato, broccoli, wheat, cauliflower, rice, cabbage,					
CC	canola, bean, soybean, corn, barley, citrus, cotton, cassava and walnut					
CC	having altered levels of ELF3 protein. ELF3 regulatory sequence can be					
CC	linked to a foreign gene and used to obtain a circadian clock-based gene					
XX	expression					
SQ	Sequence 695 AA;					
Query Match						
Best Local Similarity		100.0%;		Score 3669;		DB 3; Length 695;
Matches 695;		Conservative 0;		Mismatches 0;		Indels 0; Gaps 0;
QY	1	MKRGKDEEKILEPMPRLHVNDADKGGPRAPPKKNKALYEQLSIPSRFGDGTNRSN	60			
Db	1	MKRGKDEEKILEPMPRLHVNDADKGGPRAPPKKNKALYEQLSIPSRFGDGTNRSN	60			
QY	61	NTSTLVHPGSPQCGVERNLVQHLDSAAANQATEKFVSQMFENVRSSAQHDQKAV	120			
Db	61	NTSTLVHPGSPQCGVERNLVQHLDSAAANQATEKFVSQMFENVRSSAQHDQKAV	120			
QY	121	REEDFAPVYINRRSQSHGRTKSGIEKEKHTPMVAPSSHHHSIRFQEVNQTGSKQNVCL	180			

Db	121	REEDFAPVYINRRSQSHGRTKSGIEKEKHTPMVAPSSHHHSIRFQEVNQTGSKQNVCL	180			
QY	181	ATCSKPEVRDOVKANARSGGFVLSLJVSVTTEEIDLEKSGASSHDRVNDYNASLQESGRNRL	240			
Db	181	ATCSKPEVRDOVKANARSGGFVLSLJVSVTTEEIDLEKSGASSHDRVNDYNASLQESGRNRL	240			
QY	241	YRDGGKTRLKDTONGAESHLATENHSGEGHSPEDINDREYSKSRACASLQIQNEASD	300			
Db	241	YRDGGKTRLKDTONGAESHLATENHSGEGHSPEDINDREYSKSRACASLQIQNEASD	300			
QY	301	DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQORVFAVQLFELHRLIKVQKL	360			
Db	301	DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQORVFAVQLFELHRLIKVQKL	360			
QY	361	IAASPDLLIDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVVKORGDEKTDQHKMES	420			
Db	361	IAASPDLLIDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVVKORGDEKTDQHKMES	420			
QY	421	SAENVVGRLSNQGHQHSQSNYMPFANNPPASPAFNGYCFPPQPPPPSGNHQQLIPVMSPSSE	480			
Db	421	SAENVVGRLSNQGHQHSQSNYMPFANNPPASPAFNGYCFPPQPPPPSGNHQQLIPVMSPSSE	480			
QY	481	GLYKHPGMAHTCHYGGYGYHYMPTMVMVMPQHPGMPGPPPGNGYFPYPYGMPTTMNYPY	540			
Db	481	GLYKHPGMAHTCHYGGYGYHYMPTMVMVMPQHPGMPGPPPGNGYFPYPYGMPTTMNYPY	540			
QY	541	CSSQOQOQOQPNQMNQFQHPGNLQNTQOQOQSDNEPAPQOQOQPTKSYPRARKSRQGS	600			
Db	541	CSSQOQOQOQPNQMNQFQHPGNLQNTQOQOQSDNEPAPQOQOQPTKSYPRARKSRQGS	600			
QY	601	TGSSPSGQGISGSK3FRPFAAVDEDSNINNAPEQTMTTTTTTTTRTTVTQTTTRDGGGVTR	660			
Db	601	TGSSPSGQGISGSK3FRPFAAVDEDSNINNAPEQTMTTTTTTTTRTTVTQTTTRDGGGVTR	660			
QY	661	VIKVPHNAKLASENAAARIFQSIOEERKRYDSSKP	695			
Db	661	VIKVPHNAKLASENAAARIFQSIOEERKRYDSSKP	695			
RESULT 2						
AAO15964	AAO15964	standard; protein; 695 AA.				
ID	AAO15964					
XX	AAO15964;					
AC	AAO15964;					
XX	07-FEB-2003	(first entry)				
DT	07-FEB-2003	(first entry)				
XX	Arabidopsis thaliana nematode feeding site-specific protein #2.					
DE	Arabidopsis thaliana nematode feeding site-specific protein #2.					
XX	Nematode feeding site-specific; cyst nematode inducible transcription;					
KW	regulatory DNA sequence; root knot inducible transcription;					
KW	nematode-resistant plant.					
OS	Arabidopsis thaliana.					
XX	US6448471-B1.					
FN	US6448471-B1.					
XX	10-SEP-2002.					
PD	10-SEP-2002.					
XX	21-JAN-1999;	99US-00234827.				
PF	21-JAN-1999;	99US-00234827.				
XX	22-JAN-1998;	98US-0072142P.				
FR	22-JAN-1998;	98US-0072142P.				
XX	(PUZI/) PUZIO P S.					
FA	(GRUN/) GRUNDLER F M W.					
XX	Puzio PS, Grundler FMW;					
PI	Puzio PS, Grundler FMW;					
XX	WPI; 2003-038247/03.					
DR	N-ESDB; AAL50923.					
XX	New isolated regulatory DNA sequences obtained from Arabidopsis thaliana,					
PT	useful for promoting root knot and cyst nematode-inducible transcription					

Db	121	REEDFAPVYINRRSQSHGRTKSGIEKEKHTPMVAPSSHHHSIRFQEVNQTGSKQNVCL	180
QY	181	ATCSKPEVRDQVKANARGGFFVSLDVSVTEDIDLEKSASSHSDRVNDYNASIRQSRNRL	240
Db	181	ATCSKPEVRDQVKANARGGFFVSLDVSVTEDIDLEKSASSHSDRVNDYNASIRQSRNRL	240
QY	241	YRDGGKTRLDKTDNGAESHLATENHSORGHGSPEDIDNDREYSKSRACASLQOINEEASD	300
Db	241	YRDGGKTRLDKTDNGAESHLATENHSORGHGSPEDIDNDREYSKSRACASLQOINEEASD	300
QY	301	DVSDDSMVDSISSIDVSPDDVVGILGQKRFWARKAIANQQRVFAVQLFELHRLIKVQKL	360
Db	301	DVSDDSMVDSISSIDVSPDDVVGILGQKRFWARKAIANQQRVFAVQLFELHRLIKVQKL	360
QY	361	IAASPDLLDLDEISFLGKVSASYPVKLLPSFLVKPPLPHVTVVQKRGDSEKTDQHKMS	420
Db	361	IAASPDLLDLDEISFLGKVSASYPVKLLPSFLVKPPLPHVTVVQKRGDSEKTDQHKMS	420
QY	421	SAENVVGRLSNQHQQSNYMFANPPASAPNGYCPFPQPPSGNHQOOLIPVMSPE	480
Db	421	SAENVVGRLSNQHQQSNYMFANPPASAPNGYCPFPQPPSGNHQOOLIPVMSPE	480
QY	481	GLIYKPHGMAHTGHYGYGYHYMPTVMVMPQYHFGMGFPFPPGNGYFPPYGNMPTIMNY	540
Db	481	GLIYKPHGMAHTGHYGYGYHYMPTVMVMPQYHFGMGFPFPPGNGYFPPYGNMPTIMNY	540
QY	541	CSSQQQQQQQNEQWQNGHGNLONTQQOQORSNEPAPQOQQPTKSYPRARKSRQS	600
Db	541	CSSQQQQQQQNEQWQNGHGNLONTQQOQORSNEPAPQOQQPTKSYPRARKSRQS	600
QY	601	TGSSPSGPGISGSKSFRPFAAVIDSDSNINNAPEOTMTTTTTTTTTTTTQTTRDGGVTR	660
Db	601	TGSSPSGPGISGSKSFRPFAAVIDSDSNINNAPEOTMTTTTTTTTTTTTQTTRDGGVTR	660
QY	661	VIKVVPNAKLASENAARIFQSIQERKRYDSSKP	695
Db	661	VIKVVPNAKLASENAARIFQSIQERKRYDSSKP	695
RESULT 2			
AAO15964			
ID	AAO15964	standard; protein; 695 AA.	
XX	AAO15964;		
XX	07-FEB-2003	(first entry)	
DT	Arabidopsis thaliana nematode feeding site-specific protein #2.		
DE	Nematode feeding site-specific; cyst nematode inducible transcription;		
XX	regulatory DNA sequence; root knot inducible transcription;		
KW	nematode-resistant plant.		
OS	Arabidopsis thaliana.		
XX	US6448471-B1.		
XX	10-SEP-2002.		
XX	21-JAN-1999;	99US-00234827.	
XX	22-JAN-1998;	98US-0072142P.	
XX	(PUZL/) PUZIO P S.		
XX	(GRUN/) GRUNDLER F M W.		
PI	Puzio PS, Grundler FMW;		
XX	WPI; 2003-038247/03.		
DR	N-PSDB; AAL50923.		
XX	New isolated regulatory DNA sequences obtained from Arabidopsis thaliana,		
PT	useful for promoting root knot and cyst nematode-inducible transcription		

```
PT of an associated DNA sequence.
XX
PS Disclosure; Col 39-42; 29pp; English.
XX
CC The invention comprises an Arabidopsis thaliana nematode feeding site-specific protein #1.
CC specific regulatory DNA sequence that promotes root knot and cyst
CC nematode inducible transcription. The Arabidopsis thaliana DNA sequence
CC of the invention is useful in promoting root knot and cyst nematode
CC inducible transcription of an associated DNA sequence when re-introduced
CC into a plant. The Arabidopsis thaliana DNA sequence is useful in making
CC plants that are resistant, or at least less susceptible to plant
CC parasitic nematodes. The present amino acid sequence represents an
CC Arabidopsis thaliana nematode feeding site-specific protein
XX
XX Sequence 695 AA;
Query Match 90.7%; Score 3327; DB 6; Length 695;
Best Local Similarity 91.2%; Pred. No. 2.1e-261;
Matches 649; Conservative 4; Mismatches 25; Indels 34; Gaps 3;
QY 1 MKRGDEEKILEPMPRLHVNDADKGGPRAPRNKMALEYQLSIPSRFGDGTWNSRSN 60
Db 1 MKRGDEEKILEPMPRLHVNDADKGGPRAPRNKMALEYQLSIPSRFGDGTWNSRSN 60
QY 61 NTSTLVHPGSPQPCGVERNLSVQHLDSAAQAATEKFVQSMFNVSRSSAQHDQRKMV 120
Db 61 NTSTLVHPGSPQPCGVERNLSVQHLDSAAQAATEKFVQSMFNVSRSSAQHDQRKMV 120
QY 121 REEDFAPVYINRSRSHGRTKSGIEKEKHTPMVAPSSHHISIRFOEVNQTGSKQNVCL 180
Db 121 REEDFAPVYINRSRSHGRTKSGIEKEKHTPMVAPSSHHISIRFOEVNQTGSKQNVCL 180
QY 181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSSASHDRVNDYNASLRQSRNRL 240
Db 181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSSASHDRVNDYNASLRQSRNRL 240
QY 241 YRDGKTRKLTNDNGAESHLATENHSQEGHSPEDINDREYKSRACASLQGINEEASD 300
Db 241 YRDGKTRKLTNDNGAESHLATENHSQEGHSPEDINDREYKSRACASLQGINEEASD 300
QY 301 DVSDSMVDSISSIDVSPDDVVGILGQRFWRARAKAIANQORVAVQLFELHRLIKVQKL 360
Db 301 DVSDSMVDSISSIDVSPDDVVGILGQRFWRARAKAIANQORVAVQLFELHRLIKVQKL 360
QY 361 IAASPDLLIDELISFLGKYSAPYVKLLPSSEFLVKPPLPHVVKQGDSEKTDQHKMES 420
Db 361 IAASPDLLIDELISFLGKYSAPYVKLLPSSEFLVKPPLPHVVKQGDSEKTDQHKMES 420
QY 421 SAENVVGRLSNOGHQSQSNMYPFANNPPASPAPNGYCFPPPPGPNHQ----- 469
Db 421 SAENVVGRLSNOGHQSQSNMYPFANNPPASPAPNGYCFPPPPGPNHQ----- 469
QY 470 --QWLI PWMSRSEGLI----YKHPGMAHTGHYGYGHYMPMPMPQYHFGMGFPBP 523
Db 470 --QWLI PWMSRSEGLI----YKHPGMAHTGHYGYGHYMPMPMPQYHFGMGFPBP 523
QY 524 NGYFFPYGMPTIMNPYCQQQQQQQNEQNMNQFCHPCNLQNTQQQQORSNEPAPQQ 583
Db 524 NGYFFPYGMPTIMNPYCQQQQQQQNEQNMNQFCHPCNLQNTQQQQORSNEPAPQQ 583
QY 584 QOPTKSYPRARKSRGSGTSSPGQIGSGSKSFRPFAAVDSDSINNAPETMTTTTT 643
Db 584 QOPTKSYPRARKSRGSGTSSPGQIGSGSKSFRPFAAVDSDSINNAPETMTTTTT 643
QY 644 TRTTVTOTRDGGVTRVTKVVPNNAKLASNAARIFQSIQERKRYDSSKP 695
Db 644 TRTTVTOTRDGGVTRVTKVVPNNAKLASNAARIFQSIQERKRYDSSKP 695
RESULT 3
AAO15963
ID AAO15963 standard; protein; 695 AA.
XX
```


DR WPI; 2003-250078/25.
DR N-PSDB; ACC61147.
XX New isolated protein complexes useful for diagnosing a disease or
PT disorder, or as a target for an active agent of a pharmaceutical,
PT preferably a drug target in the treatment or prevention of disease or
PT disorder.
XX
PS Disclosure; SEQ ID NO 1075; 17pp + Sequence Listing; English.
XX
XX The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM
XX
XX Sequence 802 AA;
SQ
Query Match 5.3%; Score 193.5; DB 6; Length 802;
Best Local Similarity 20.2%; Pred. No. 1.5e-06;
Matches 128; Conservative 73; Mismatches 216; Indels 217; Gaps 29;
QY 81 LSVQHLDSAAQAQTEKFVSQMFENVRSSAQHDQK-----VREEDFAVPV 130
DB 253 LSVPPHDSITSTASATGARSNDQRPPLSDAQRMRFPADLSKPIITEEHDYAA 312
QY 131 YINRSRSGHGRK-----SGIEKEKHTPMVAPSPSHHSIRPQEVN 170
DB 313 YL-----HGENKITEMHNIPPKSLFLIGNPLKNVSKEDLPRIFPYGH-----IM 358
QY 171 QTSKQNVCIATCSKE-VRDQVKANARSGGFVSLDVSVEIDLEKSASS-----H 222
DB 359 QINIKNAFGFIQDNQPSVDAIECE-----SQEMNFGKLLILEVSSSNARPQFDHG 410
QY 223 DRVNDYNASLRQSRNRLYRDGKTRIKDQDNG-----AEGLHAT 262
DB 411 DHGTNSSTTSSAKRPFTQESD--MYNDNAGYKKSRHTVSCNIFVKRTADRYAI 468
QY 263 E--NHSGEGHSPEDINDREYSKRACASLOQINEASDDVSDSDSISIDVSPDD 320
DB 469 EVENRPDGTG---LETDMIFLKRMELG-KLINDAAYNGVGVVLVKNTHNVDV--- 519
QY 321 VVGLGQKRWARKAIAANQORFVAVQLFELHRLIKVOKLIAASPDLLDLDEISPLGKVA 380
DB 520 -----QTFYKGSQGET-----KFDYISISAD---DAVALFNKIK- 551
QY 381 KSPYVKLLPSEFLVAKPLPHVVKQDSEKTDQHKWSSSAENVVGRLSNQGHQHSQSNY 440
DB 552 -----NNRNSRPDIYRANS-----HQQNIY 572
QY 441 -----MPFANNPPASPAP-----NGYCFPP-----QPPSGNHQOQMLIPVMSPEGLIYKPH 487
DB 573 GAPPPLVPNGFVAGFPPTQNYQYSWMPFPQQOQQOQPYGNYG---MPPPSHDQG--YGSQ 627
QY 488 PGMAHTGHYGYHYHPTMVMPOYHPGMGF-----PPGNGYFPP----- 529
DB 628 PPIPMNOSYGRYQTSIPPPP---PQOIPOGYGRYQAGPPQP-----PPSQTPMQOQQLS 680
QY 530 --YGMMPITIMPYCSSOQQOQQOQNEQMNQFQGFNLQNTQQOQOQSDNEPAPQOQOQPT 587
DB 681 AIQNLIPNVVSNLLSMAQOQQOQOQHAQOOLV---GLIQSMQO-----APQOQOQOL 729
QY 588 KSYPRARKSGSTGSSPSGPGQIGSGSKFRPPFA 621
DB 730 GGYSSMNSSPPPMSTNYNG-QNISAKPSAPPMS 762

RESULT 7
ID ABB60944 standard; protein; 1503 AA.
XX ABB60944;
AC ABB60944;
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 9624.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX N-PSDB; ABL05047.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
PT
XX Disclosure; SEQ ID NO 9624; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1503 AA;
Query Match 5.2%; Score 192; DB 4; Length 1503;
Best Local Similarity 22.9%; Pred. No. 4.7e-06;
Matches 151; Conservative 74; Mismatches 244; Indels 190; Gaps 35;
QY 4 GKDEEKILEMFP-----RLHVNDADKGGPRAPPRNKMALYEQLSIPSRFGDHTGNSR 58
DB 398 GMDGNRME--LPQSYDYRVHQYD-----RFHPLNQ-----RQMS-----NVGDY-----ER 437
QY 59 SNTTSTLVHPCSPSQPCGVNRLSVQHLDSAAQAQTEKFVSQMFENVRSSAQHDQK 118
DB 438 SNR-----MPLDLDFKREPHSESSFLAPNELSLGQLSVVEEPLKS-----EDEN 483
QY 119 MVREEDFAVPVYINRSRSGHGRKTSKIEKHTPMVAPSSHHSIRFQVNTGSKQNV 178
DB 484 ISNREMAKIAV-LDIKDDKALHSTGDGSEKEDT-----LEEVKLEGLHEEL 530
QY 179 CL--ATCSKPEVRDQVKANARSGGFVSLDVSVEETIDLEKSASSHSDRVNDYNASLQES 236
DB 531 KLDEESLVKPESKLDSVSCINGKF-----EDRSEMEVENKPLGPVK-REAKAELINQM 583
QY 237 RNLRYDGGKTRIKDQDNGAESHLATENHSQEGHSGSPEDINDREYSK-SRACASLOQIN 295

Db 584 GNKLKNDGSTLVAK---SSAESPLKEDVPPNLRARSELNVSTGTDLSAVSKQNSLKRRK 640
QY 296 EASDDVDSD--DSMVD-----SISIDVSPDDVVGILGQKFWFARKAIAQORV 343
Db 641 SSATSDSDKHSNDPCDTECKSEDAEKSETDAEKDD----- 677
QY 344 FAVQIFELHRLIKVQKLIASPDLLDETS--FLGKVSASYP-----VKLLPSEFLVK 396
Db 678 ---SIVKNIKILQEPVEKSE--CEEISKVVGGAAASAPKVPDPAVVLLANLSMD 732
QY 397 PPLPHVVVKQSG-----DSEKTDHKNMSSAENVGRLS---NOGHH 435
Db 733 -----KDRGKQARHRYRQRNRRSKRSLDNDQHESDDKQHAVSKRLAPLJLDDFD 784
QY 436 QOSNYMPFANNPPASGPNAGYCFPPQPPPSGNHQWLIPVMSPEGLIYKHPGMAH--- 492
Db 785 HNGNRVGFNG-----LNG-MLTDPDAEYSGERS-----YPSHEMLAHFDE 823
QY 493 --TGHYG-GYGYHYMPTMVMPOYH--PGMGFPFGNGYFPYFGMMPTIMNYPYCSQQQ 547
Db 824 LETAHGDSMVIYDPTKVVVPMVWQRPNGGF-----QGQOPVQMQ--FSQQQP 870
QY 548 QOOPNEQMNQFCH-----PGNLQNTQQOQOORSDNEPAPQOQO-QPTKSYPRARKSRQS 600
Db 871 QOQOQOQOQOQO-CHMQOQPPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 928

RESULT 8
ID ABG93140 standard; protein; 738 AA.
XX AC ABG93140;
XX DT 21-NOV-2002 (first entry)
XX DE S. cerevisiae BAX-associated protein fragment SEQ ID 238.
XX KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW neurodegeneration; cell death.
XX OS Saccharomyces cerevisiae.
XX PN WO200264766-A2.
XX PD 22-AUG-2002.
XX PF 21-DEC-2001; 2001WO-EP015398.
XX PR 22-DEC-2000; 2000EP-00870318.
XX PR 04-JAN-2001; 2001EP-00870002.
XX PR 09-JAN-2001; 2001EP-00870003.
XX PA (JANC) JANSEN PHARM NV.
XX PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX DR WPI; 2002-667002/71.
XX DR N-PSDB; ABQ76406.
XX PT New isolated nucleic acid representing a synthetic BAX-gene, useful as
XX PT medicament for treating, preventing and/or alleviating yeast or fungal
XX PT infections or proliferative disorders, or for preventing apoptosis in
XX PT certain diseases.
XX PS Claim 36; Fig 1; 344pp; English.
XX CC This invention describes a novel nucleic acid representing a synthetic
XX CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
XX CC resistant yeast or fungi, identifying, or obtaining and identifying
XX CC Candida spp. sequences that are differentially expressed in a pathway
XX CC eventually leading to programmed cell death or identifying inhibitors or

CC inhibitor sequences of Bax-induced cell death. The products of the
CC invention have cytostatic, fungicide; immunosuppressive, virucide and
CC vasotropic activity and can be used in vaccines or for gene therapy. The
CC isolated nucleic acids, polypeptides, pharmaceutical compositions, or in
CC antisense molecules and antibodies are useful as medicaments or in
CC preparing a medicament for treating, preventing and/or alleviating
CC diseases associated with yeast or fungi or proliferative disorders, such
CC as cancer, or for preventing apoptosis in certain diseases. The compounds
CC or polypeptides, or the genetically modified organism are useful for
CC preparing a medicament for modifying the endogenic flora of humans and
CC other mammals. The vaccine is useful for immunising against yeast or
CC fungal infections. Apoptosis-related diseases include autoimmune disease,
CC ischaemia, diseases related with viral infections or neurodegenerations.
CC This sequence represents a polypeptide associated with the Bax gene
CC described in the disclosure of the invention
XX
SQ Sequence 738 AA;
Query Match 5.2%; Score 189.5; DB 5; Length 738;
Best Local Similarity 17.5%; Pred. No. 2.8e-06;
Matches 128; Conservative 99; Mismatches 248; Indels 257; Gaps 29;
QY 19 HVN-DADGGPAPPNNKALYEQLSIPSQRPDHGT-MNRSNNTSTLVHPGPSQPG 76
Db 142 HVNLDKGHVSPKPSVNTTWSAAAASVSDTK---HDVPQDSNDNNNEELEAQGQAQEK 197
QY 77 VERNLSVCHLDSSAANOATEKPVSQMSPMENVSSAQHDQRKMWREEDFAVPVINSRR 136
Db 198 QKEQEEQQQEGHNNKEHKQIEQPS-----LSSKK 229
QY 137 SQSHGRTKSGIEKXHTPMVAPSSHHSIRFOVNTQSGKNVCLATCSKPEVRDQVKANA 196
Db 230 TTS--RTSASQPKMSWAAIATPKPAVKKTE---SPLNVAEL---KKEISDIK--- 277
QY 197 RSGGFVSLDVSTVEIDLEKXASSHDRVNDYNASIROESRNLRYDGGKTLKTDNGA 256
Db 278 -----DDKSEASEEKVNEOTSQOEPE----- 302
QY 257 ESHLATENHSQBGHSGSPDINDREYSKRACASQQTNEEASDDVSDSDMVDSTSSIDV 316
Db 303 -----TAPSENEEDRVPEVDGEEVQEEAE---KKEQVKEEBEQTAEELEQEOQDNVA--- 350
QY 317 SPDDVVGILGQKRFWRARKAIAQORVFAVQLFELHRLIKVQKLIASPDLLDLSIFLG 376
Db 351 APEEEVTVVERK-----VEISAVISEPPE---DQANTVP 381
QY 377 KVSAS-----YPVKLLPSEFLVKPPLPHVVVKRGDSEKTDHKNMSSAENVV 426
Db 382 QPQQSQSQP 440
QY 427 GRLS-----NOGHHQOQSNM-----PFANNPPASAPNGY---CFPPQPPPS 465
Db 441 EQLSQNYTQQOQ 500
QY 466 GNH---QOWLI-----PVMSPSSEGLIY----- 484
Db 501 GNNVAAQOQYVMQNFPGVSGYPMFDSQGYAYGQOYQQLAQNNNAQTSGNANYNFQOQY 560
QY 485 -----KPHPGMAHT-----GHYGG-----YYGHY--MPTPMVMPQVHP 515
Db 561 QAGANTAAANLTSAAAAAASPATAHAQPOQOQPYGSGFMPYAHFYQSQSPFPYQPGY 620
QY 516 GMGEP--PPGNGYPPYGMWMTIMNYPYCSQOQOQOQOQOQOQOQOQOQOQOQOQOQO 573
Db 621 AQQYPYQLPKNNY-----NYOTQNGEQOQSNQGVQAQHSQSQOQSQOQOQOQOQO 669
QY 574 -----SDNEPA-PQOQOQPTKSYPRARKSRQSGTSGSPSPGQSGSKSRFPF 620
Db 670 QPQGPQPVEVQMNGQFVNPQQQMQFQCYQFQOQOQAAAAAQAQGV-----PY 722
QY 621 AAVDEDSNINNA 632
Db 723 GYNGYDYNKNS 734

XX	RESULT 9	
XX	ABP73921	
ID	ABP73921 standard; protein; 1130 AA.	
XX		
AC	ABP73921;	
XX		
DT	30-JAN-2003 (first entry)	
XX		
DE	Candida albicans essential protein SEQ ID NO 7758.	
XX		
KW	Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;	
KW	signal transduction; DNA replication; cell division; growth;	
KW	proliferation; Candida albicans; fungicide; antifungal.	
XX		
OS	Candida albicans.	
XX		
PN	WO200253728-A2.	
XX		
PD	11-JUL-2002.	
XX		
PF	26-DEC-2001; 2001WO-US049486.	
XX		
PR	29-DEC-2000; 2000US-0259128P.	
PR	20-FEB-2001; 2001US-00792024.	
XX	22-AUG-2001; 2001US-0314050P.	
XX		
PA	(ELIT-) ELITRA PHARM INC.	
XX		
PI	Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;	
XX		
DR	WPI; 2002-566694/60.	
DR	N-PSDB; AB232471.	
XX		
PT	Constructing strains for identifying gene products as effective targets	
PT	for therapeutic intervention, by inactivating in the strain one allele of	
PT	a gene and placing other allele of the gene under conditional expression.	
XX		
PS	Claim 44; SEQ ID NO 7758; 167pp + Sequence Listing; English.	
XX		
CC	The invention relates to constructing (M1) a strain of diploid fungal	
CC	cells in which both alleles of a gene are modified, comprising modifying	
CC	one allele by insertion or replacement by a cassette having an	
CC	expressible selectable marker and modifying other allele by	
CC	recombination, of a promoter replacement fragment with a heterologous	
CC	promoter, so that expression of the second allele is regulated by the	
CC	promoter. (M1) is useful for constructing a strain of diploid fungal	
CC	cells in which both alleles of a gene are modified. The diploid fungal	
CC	cells having both alleles modified are useful for identifying a gene that	
CC	is essential to the survival or growth of a fungus, a gene that	
CC	contributes to the virulence and/or pathogenicity of a fungus, a gene	
CC	that contributes to the resistance of a diploid fungus to an antifungal	
CC	agent, an antifungal agent that inhibits the growth of a diploid fungus	
CC	and for identifying a therapeutic agent for treatment of a mammalian	
CC	disease. (M1) is useful for identifying a compound which modulates the	
CC	activity of a gene product, preferably enzymatic activity, carbon	
CC	compound catabolism, biosynthetic, transporter, transcriptional,	
CC	translational, signal transduction, DNA replication and cell division	
CC	activity. The method is useful for identifying a compound having the	
CC	ability to inhibit growth or proliferation of C. albicans cells and for	
CC	treating infection by C. albicans. The present sequence is that of an	
CC	essential Candida albicans protein used in the method of the invention.	
CC	Note: The sequence data for this patent is not represented in the printed	
CC	specification but is based on sequence information supplied to Derwent by	
CC	the European Patent Office	
XX		
SQ	Sequence 1130 AA;	
XX		
Query Match	5.2%; Score 189.5; DB 5; Length 1130;	
Best Local Similarity	21.0%; Pred. No. 5.1e-06;	
Matches 130; Conservative	83; Mismatches 228; Indels 179; Gaps 32;	

155

MVAPSSHSIRFOEVNQTGSK---QNVCLATCSKEPEVRDQVKANARSGGFVISLDSVTE

211

6

MSAPPSIPTITPGTSSINTRRVARACLSCKEK-----KIKNGEPMTTITTAAD--GTN

58

212

BIDLEKSAS-SHDRVNDYNASLRQESRNLRYDGGKTLKOTD---NGAESHLATENHSQ

267

59

RIIPEKTRTCNCRFLGIPCVFVQSNR-----GGKRKRSTEHOLSEATAPATH--

110

268

EGHGSPEIDIDNRE--YSKSRACASLQOINEEASDDVDSDMSVDSISSIDVSPDVVGI

324

111

-----PNOLGHOHQLPPIKTKTSPSIHNL-----DSGVV--IPSI-----

144

325

LQQRKFMARAKAIAANQORVAVQLFELHRLIKVKLIAASPDLLDLDEISFLGKVSAKYP

384

145

-----NQQERHQDGYRNEELASTERV---SSYILD-----GNNSTK--

178

385

VKKLLPSEFLVKPPLPHVVVQKRGSEKTDQHKMESSA-----ENVVGRLS--

430

179

-----LPS-----PVLTHTTFSRRDEGQRFNNQLRRSSTTTAGFSLPQPQELGRRSVY

229

431

NOGHHQSNYMPF-----ANPPASPAPNGYCFPPPPP-----PSGNHQQWLI

473

230

SPGPSIKSQLPILTSSSTTTSSVQSPPPPP-----PPQPRGMGIQFHEASHQGI

283

474

PVMSPEGLIYKPHPCMAHTGH-YGYGYGHYMTPT-----MVMPOYHPGMGFP

521

284

PPPPP-----PPPPPPPHDLHHQYGYGPPPPPPPPPPPLHHHHHHYPPSHHPG

537

522

PGNGYFPYGMPTIMNPYCSQQQQOQPNQMGHFG--NLQNTQQOQOOR-----

573

338

PPPGPPGPPPPPPPPPHYHOSAPESPPEKFKHR--HPKYFKONKKGHRHHPHLH

395

574

SDNEPAPOQ-----QQOPTKSPRARKSRQGSTGSSPGPQ-----GISGSKSFR--

618

396

HHHPPPHHHHHPDFQHHNGQDDIKSKHTDENSQNSKSPRWLDNMSVSSDSSRSV

455

619

--PFAAVDBSDNINNAPEQTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTT

676

456

SSPQSHNDKQSVSNHKSSTMSCT--NSVTQT-----LDRLPKIVSTQONNLPTS

506

677

ARIFQS--IQEERKRYDSSK

694

507

KILNSPFTESLAKYDLPK

526

RESULT 10

AB005434

ID

ABJ05434 standard; protein; 1138 AA.

XX

AC

ABJ05434;

XX

DT

14-NOV-2002 (first entry)

XX

DE

Human breast cancer associated polypeptide SEQ ID NO: 194.

XX

KW

Human; breast specific gene; breast specific protein; breast cancer;

KW

gene therapy; cytostatic.

XX

OS

Homo sapiens.

XX

PN

WO200264611-A1.

XX

PD

22-AUG-2002.

XX

PF

12-FEB-2002; 2002WO-US004197.

XX

PR

13-FEB-2001; 2001US-0268292P.

XX

PA

(DIAD-) DIADEXUS INC.

XX

PI

Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;

PI

Sun Y, Liu C;

XX

DR	WPI; 2002-657582/70.	
XX	New breast specific nucleic acids and proteins, useful for identifying,	
PT	diagnosing, monitoring, staging, imaging, and treating breast cancer and	
PT	non-cancerous disease states in breast tissue, and in gene therapy.	
XX	Claim 11; Page 293-298; 367pp; English.	
PS	The present invention provides human breast specific coding sequences and	
CC	proteins. These can be used in the diagnosis and treatment of breast	
CC	cancer and non-cancerous diseases of the breast. The present sequence is	
CC	a polypeptide of the invention	
XX	Sequence 1138 AA;	
SQ	Query Match 5.1%; Score 186.5; DB 5; Length 1138;	
	Best Local Similarity 19.3%; Pred. No. 9e-06;	
	Matches 137; Conservative 90; Mismatches 235; Indels 249; Gaps 30;	
QY	125 DFAVP-----VYINRRSSQSHGRKSGIEKEKHTPMVAPSSH-----162	
Db	3 DFAAPAAAAGSSICINSSINSSLGAGIGVNTPTNSTPAAPSSNHPAAGCGGSGPGG 62	
QY	163 -SIRFQEVNQTGSKQNVCIATGSKPEVRDOVKANARSGFVSLDVSVTEIDLEKSASS 221	
Db	63 GSAAVPKHSTVVERLFQRLEGCRHRHVNCENR-----YQQAQVEQLELER---- 107	
QY	222 HRVNDYNASLRQESRNLRYRGGKTRLDXTDNGAESHILATE-NHS-----QEG 269	
Db	108 RUTVSLYQRTLRQAKGAGTGKQHPKQODAEASAEQRNHTLIMLQTVRKKEUG 167	
QY	270 HGSPEIDINDREYSKRAC-ASLQIQINEASDDVSDSMVDTSISI-----DVSPDQV 321	
Db	168 ARSP--LNGDQQ--NGACDGNFSPSKIRKDIS--AGMEAINNLNPMPLPSASP--- 217	
QY	322 VCILGQKRFWRARKALANQORFVAVQLFELHRL-----IKVKQLIAASPDLLDLEISF 376	
Db	218 -----LHQLDLKPSLPLQNSGTHTPG-LLEDLSKNG 247	
QY	377 KVSASYPVK-----KLAPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESS----- 421	
Db	248 RLPEIKLPVNGCSDLEDSTFIQSKDLKQEPDLDPTCIDTSETSLNQNKLFSLNLNDQ 307	
QY	422 -----AENVVGRSLNQGHQOS-----NYMPFA 444	
Db	308 EQOELIDELANTVPEDDIQDLFNEDEFEKKEFEFSQPATETPLSQESASVKSDPSHPFA 367	
QY	445 N-----NPPASPAPEGVCFPP-----QPPSGNHQOQLIPVMSPEGLIYKP 486	
Db	368 HVSMGSPQARPSSG---PPFSTVSTATSLPSVASTPAAPN-----PASSPANCAVQSP 418	
QY	487 H-PGMAHTGHYGYYHYMPTPMWMPQVHPGMPFPFGNGYF-----527	
Db	419 QTPNQAH-----PGQAPRPGNGYLLNPAAVTVAGSAGPVAV 457	
QY	528 PPGMPTIMPYCSQQQQQQQPPNEQMNQFHGPNLQNTQQQQQSDNEPAPOQQQQPT 587	
Db	458 PSSDMSPAELKQMAAQOQOQAKLMQOQKQ-----QQQQQQQQQQQQQQQQQQQQ 510	
QY	588 KSYPRARKSRQSTGSSPSGPG--ISGSKSRPRFAAVIDEDSNINN-APEQMTTTTTT 644	
Db	511 HSNQTSNNSPLG---PPSSPYCAAFATKPNSPMWYPOAFNNQNPVPPMANNLQKTTM 566	
QY	645 RTVTQTDTGGVTRVIKVPVHNKLAASENAARIFQSIQERKBYDSSKP 695	
Db	567 NNYLPQNEMN-----MINQPPNNNLGTSLNK-----QHNILTYGNTPK 604	
RESULT 11		
ID	ABBS8529	
XX	ABBS8529 standard; protein; 771 AA.	
AC	ABBS8529;	
XX	DT (first entry)	
DE	Drosophila melanogaster polypeptide SEQ ID NO 2379.	
XX	Drosophila; developmental biology; cell signalling; insecticide;	
KW	pharmaceutical.	
XX	Drosophila melanogaster.	
OS	WO200171042-A2.	
PN	27-SEP-2001.	
XX	23-MAR-2001; 2001WO-US009231.	
PF	23-MAR-2000; 2000US-0191637P.	
XX	11-JUL-2000; 2000US-00614150.	
PR	(PEKE) PE CORP NY.	
XX	Venter JC, Adams M, Li PWD, Myers EW;	
PI	WPI; 2001-656860/75.	
XX	N-PSDB; ABL02632.	
DR	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	genes from Drosophila and for elucidating cell signaling and cell-cell	
PT	interactions.	
XX	Disclosure; SEQ ID NO 2379; 21pp + Sequence Listing; English.	
PS	The invention relates to an isolated nucleic acid detection reagent	
CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
CC	useful in developmental biology and in elucidating cell signalling and	
CC	cell-cell interactions in higher eukaryotes for the development of	
CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	
CC	sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-	
CC	ABBS7202). The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX	Sequence 771 AA;	
SQ	Query Match 5.1%; Score 186; DB 4; Length 771;	
	Best Local Similarity 22.8%; Pred. No. 5.8e-06;	
	Matches 109; Conservative 56; Mismatches 174; Indels 140; Gaps 22;	
QY	265 HSQEGHSGPEDINDREYSKSRACASLQIQINEASDDVSDSMVDTSISIYSPDDVVG 324	
Db	54 HRQGG-----ESEGFAKRAATESLVKKLKERDEL--DSLITAITNGAHPKCVTI 102	
QY	325 LGQKRFWRARKAIANQQ---RVFAVQLF---ELHR-LIKVKQLIAASPDLLDLEISF--- 374	
Db	103 ---QRTLDRLQVAGKRGKGFPHVIVARIWRPDLHKNELKHVKYCAFAFDLKCDSCVCP 159	
QY	375 -----LGKVSASYPVKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAE 423	
Db	160 HYERVVSPGIDLSGLSLQSGP-SRLVKDEYSAGPLVGSN-----DIDGNDITGTHHPT 212	
QY	424 NVVGRSLNQGHQO-OSNYMPFANNPPASPAPNGYCFP---POPSPSGNHQOQLIP----- 474	
Db	213 QMVGP-GGYGYPQGPSEYVGDA-NPMSANFTGRTPIKIEPDQGVAGSRGSMVPPPPRL 270	
QY	475 -----VMSPSGLIYK---PH--PGMAHTGHYGYYHYMPTPM 508	
Db	271 GPPPOQQQQQQPQQTPQTQQQQAQSAQAASLPPVHCPGMP-----GPMNPGPV 321	
QY	509 VNPQVHPGMPGPPGNGYFPYPPYGMPTIMNYPYCSQQQQQQQOQPPNEQMNQFHGPNLQNT 568	
Db	322 MAPPPPPQQAQNPQNGVHHTQANSPTDPASALAMQQQQQQQQQQQQQ-----Q 371	

Query Match	4.9%;	Score 180;	DB 4;	Length 1594;
Best Local Similarity	18.2;	Pred. No. 4.8e-05;		
Matches 162; Conservative	95;	Mismatches 306;	Indels 328;	Gaps 35;
QY	40	BOLGIPSRFQDGHGTMSRSNNTSLVHPGFPSSQCGVERNLVSQHLDDSAANOATEKFV	99	
	:	:	:	:
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	166	QETTVLQKRFLE--SKNKRAAKTTDKLPPDSQQHQHQHQHQHQHQHQHQHQQAQTWLA	223	
QY	100	SQMSFMENVRSAQHDQRKVREED-----FAVPVYINSRSQSHGRTKSCIEKEKH	152	
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	224	QQL-----QSSVHVYQQKFLKRPADVDVNGPDSPEPPHKLPNNNNNNSNNNNNGNANN	277	
QY	153	TPWAPSSHHSLRFQEVNQTGSKQNVCLATCSKEPEVRDQVKANARSGGFVLSLDVSVTEE	212	

RESULT 13
AAAE36097
ID AAEE36097 standard: protein: 1133 AA.

XX
AC
AAE36097;

26-JUN-2003 (first entry)

XX Human Mastermind (Mam) homologue protein, MAML2.

XX Notch signal transduction; Mastermind; Mam; Mip1; Mip30; Mip6; cancer;
KW gene therapy; cytostatic; human.
KW

XX Homo sapiens.

XX
DN
W02002102987-A2

XX
C
C
C
C
C

XX

Year	Number of cases	Percentage of cases
1990	10	10.0
1991	15	15.0
1992	20	20.0
1993	25	25.0
1994	30	30.0
1995	35	35.0
1996	40	40.0
1997	45	45.0
1998	50	50.0
1999	55	55.0
2000	60	60.0
2001	65	65.0
2002	70	70.0
2003	75	75.0
2004	80	80.0
2005	85	85.0
2006	90	90.0
2007	95	95.0
2008	100	100.0
2009	105	105.0
2010	110	110.0
2011	115	115.0
2012	120	120.0
2013	125	125.0
2014	130	130.0
2015	135	135.0
2016	140	140.0
2017	145	145.0
2018	150	150.0
2019	155	155.0
2020	160	160.0
2021	165	165.0
2022	170	170.0
2023	175	175.0
2024	180	180.0
2025	185	185.0
2026	190	190.0
2027	195	195.0
2028	200	200.0
2029	205	205.0
2030	210	210.0
2031	215	215.0
2032	220	220.0
2033	225	225.0
2034	230	230.0
2035	235	235.0
2036	240	240.0
2037	245	245.0
2038	250	250.0
2039	255	255.0
2040	260	260.0
2041	265	265.0
2042	270	270.0
2043	275	275.0
2044	280	280.0
2045	285	285.0
2046	290	290.0
2047	295	295.0
2048	300	300.0
2049	305	305.0
2050	310	310.0
2051	315	315.0
2052	320	320.0
2053	325	325.0
2054	330	330.0
2055	335	335.0
2056	340	340.0
2057	345	345.0
2058	350	350.0
2059	355	355.0
2060	360	360.0
2061	365	365.0
2062	370	370.0
2063	375	375.0
2064	380	380.0
2065	385	385.0
2066	390	390.0
2067	395	395.0
2068	400	400.0
2069	405	405.0
2070	410	410.0
2071	415	415.0
2072	420	420.0
2073	425	425.0
2074	430	430.0
2075	435	435.0
2076	440	440.0
2077	445	445.0
2078	450	450.0
2079	455	455.0
2080	460	460.0
2081	465	465.0
2082	470	470.0
2083	475	475.0
2084	480	480.0
2085	485	485.0
2086	490	490.0
2087	495	495.0
2088	500	500.0
2089	505	505.0
2090	510	510.0
2091	515	515.0
2092	520	520.0
2093	525	525.0
2094	530	530.0
2095	535	535.0
2096	540	540.0
2097	545	545.0
2098	550	550.0
2099	555	555.0
2100		

FK T8-UUN-200+, 200+00

(GEHO) GEN HOSPITAL CORP.
 Artavanis-Tsakonas S, Lake RJ;
 WPI; 2003-167509/16.
 N-PSDB; AAD54622.
 Inhibiting or agonizing Notch signal transduction in a cell for preparing
 a composition for treating or preventing cancer comprises contacting the
 cell with an antagonist or agonist of sumolation.
 Disclosure; Fig 3; 84pp; English.
 The invention relates to a method for modulating Notch signal
 transduction in a cell which comprises contacting the cell with an
 antagonist or agonist of sumolation. The invention is based on the
 interactions of Mastermind (Mam) protein with the Mip1, Mip30 and Mip6
 proteins. The method is useful for preparing a composition for treating
 or preventing a disease e.g. cancer. The invention is also used in gene
 therapy. The present sequence is human Mam homologue protein used to
 illustrate the method of the invention
 Sequence 1133 AA;
 Query Match 4.9%; Score 179.5; DB 6; Length 1133;
 Best Local Similarity 19.0%; Pred. No. 3.3e-05;
 Matches 135; Conservative 91; Mismatches 232; Indels 253; Gaps 30;
 QY 125 DFAVP-----VYNSRSOSHGRKSGIEKEKHTPMVAPSSH-----162
 DB 3 DFAAPAAANGSSICINSLSLGGAGIGVNTNTPAAPSNNHPAAGCGGGGPGG 62
 QY 163 -SIRFQEVNQTGSKONVCLATCKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSASS 221
 DB 63 GSAAVPKHSVTVERLQRTEGCRHHVNCENR-----YQQAQVEQLELER----107
 QY 222 HRYVNDYNASLRQSRNRLYRDGKTRLDKTDNGAESHILATE-NHS-----QEG 269
 DB 108 RDTVSLYQRTLEQRAKSKGAGTGKQHPSPQDAAEASAEQRNHTLIMLQETVKRKLEG 167
 QY 270 HGSPEIDNDREYSKRAC-ASLQIQINEASDDVSDSDSVDSISSI-----QEG 269
 DB 168 ARSP--LNGDQO---NGACDGNFPTSKIRKDIS--AGMEAINNLPNNMPLPSASP---217
 QY 322 VGILGQKRFWRARKAIAQORVFAVQLFELHRL-----IKVQKLIASAASPDLLDEISFLG 376
 DB 218 -----LHQDLKPSLPLQNSGTHTPG-LLEDLSKNG 247
 QY 377 KVSASYPVK-----KLLPSFLVKPPLPHVVVVKRGDSKTDQHKMESS-----421
 DB 248 RLPEIKLPVNGCDLSDSFTILQSKDLKQELDDPTCIDTSETSLNQNKLFSDINLNDQ 307
 QY 422 -----AENVGRLSNOGHQOS-----NYMPFA 444
 DB 308 EQWELIDELANTVPEDDQDLFEDNEPEEKEPEFSQATETPLSOESASVKSDDPSHPFA 367
 QY 445 N-----NPPASAPNGYCFPP-----QPPPSGNHQWLLIPVMSPEGLIYKP 486
 DB 368 HVSMGSPQARPSGG--PPFSTVSTATSLPSVASTPAAPN-----PASSPANCASVP 418
 QY 487 H-PGMAHTGCHGYGYGHYNTPTMVPQVHPGMGFFPPNGYF-----527
 DB 419 QTPNQAH-----PGQAPPRPGNYLLNPAAVTVAGSAGPVAV 457
 QY 528 PPGNMPTIMNPYCSQQQQQQQNPQNMNFGHGNLQNTQQQQQSDNEPAPQQQQPT 587
 DB 458 PSSDMSPAQLKQMAAQQQQAKLUMQXQQ-----QQQQQQQQQ-----502
 QY 588 KSYPRARKSROGSTGSSPSGGOG--ISGSKSPRPFPAVDESDNINN-APEQTMTTTTTT 644
 DB 503 QQQQHSNTSNWSPLGPPSSPFGAFTAEKPNFMYQAFNNQNPVPPMANNLQKTM 562

QY 645 RTVTQTTRDGGGVTRVIVKVPHNAXLASNAARIQSIQERKRYDSSKP 695
 DB 563 NNVLPPQNMN-----MINQPPNLTGNSLNK-----QHNLITGNTKP 600
 RESULT 14
 ADC59316
 ID ADC59316 standard; protein; 1133 AA.
 XX
 AC ADC59316;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human polypeptide #4.
 XX
 KW Human; polyglutamine disease; gene; genealogical polyglutamine disease;
 KW neotrophic; anticonvulsant.
 OS Homo sapiens.
 XX
 PN JP2002360268-A.
 XX
 PD 17-DEC-2002.
 XX
 PF 03-AUG-2001; 2001JP-00236788.
 XX
 PR 04-AUG-2000; 2000JP-00236839.
 PR 06-APR-2001; 2001JP-00108723.
 XX
 PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
 PA (DAUC) DAIICHI PHARM CO LTD.
 XX
 DR WPI; 2003-516153/49.
 DR N-PSDB; ADC59315.
 XX
 PT A genealogical line diagnostic marker for polyglutamine disease, useful
 PT in the diagnosis, prevention and/or treatment, comprises a polyglutamine
 PT related gene and its encoded polypeptide.
 XX
 PS Claim 4; SEQ ID NO 8; 72pp; Japanese.
 XX
 CC The invention discloses polyglutamine disease related genes and their
 CC encoded polypeptides. Also claimed is a recombinant vector,
 CC transformants, preparation of the polynucleotides and resultant
 CC polypeptides, diagnostic methods and a kit. The genes and encoded
 CC polypeptides are useful in the diagnosis, prevention and treatment of
 CC genealogical polyglutamine disease. The sequence presented is a
 CC polypeptide of the invention.
 XX
 SQ Sequence 1133 AA;
 Query Match 4.9%; Score 179.5; DB 7; Length 1133;
 Best Local Similarity 19.0%; Pred. No. 3.3e-05;
 Matches 135; Conservative 91; Mismatches 232; Indels 253; Gaps 30;
 QY 125 DFAVP-----VYNSRSOSHGRKSGIEKEKHTPMVAPSSH-----162
 DB 3 DFAAPAAANGSSICINSLSLGGAGIGVNTNTPAAPSNNHPAAGCGGGGPGG 62
 QY 163 -SIRFQEVNQTGSKONVCLATCKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSASS 221
 DB 63 GSAAVPKHSVTVERLQRTEGCRHHVNCENR-----YQQAQVEQLELER----107
 QY 222 HRYVNDYNASLRQSRNRLYRDGKTRLDKTDNGAESHILATE-NHS-----QEG 269
 DB 108 RDTVSLYQRTLEQRAKSKGAGTGKQHPSPQDAAEASAEQRNHTLIMLQETVKRKLEG 167
 QY 270 HGSPEIDNDREYSKRAC-ASLQIQINEASDDVSDSDSVDSISSI-----QEG 269
 DB 168 ARSP--LNGDQO---NGACDGNFPTSKIRKDIS--AGMEAINNLPNNMPLPSASP---217
 QY 322 VGILGQKRFWRARKAIAQORVFAVQLFELHRL-----IKVQKLIASAASPDLLDEISFLG 376

RESULT 18	
AAG85017	
ID	AAG85017 standard; protein; 757 AA.
XX	
AC	AAG85017;
XX	
DT	06-AUG-2003 (revised)
DT	11-SEP-2001 (first entry)

capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (ABB57737-ABB2072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 333 AA;

Query Match 4.6%; Score 168.5; DB 4; Length 333;
Best Local Similarity 28.8%; Pred. No. 4.8e-05;
Matches 69; Conservative 20; Mismatches 78; Indels 73; Gaps 14;
QY 429 LSNQGHQOSNMYFANNPASPANGYCFPPQPPSGNH-----QQWLIPYMPSEGL 482
Db 8 MNSQGFHDQOH--PQAGN---NWYFGN-FQQQQPPGNGHFDPRGNFFIDPMS----- 56
QY 483 IYKPHGMAHTGHYGGYGYHYMPTMVMQYHPGNGFPP---PGNGYFPPYGMPTMMP 539
Db 57 --KPEFG-----FPGPMHNGPAMGQQQPPGQ-FPPMQAPGSGNEPVDLGVLSAAI 107
QY 540 YCSSQQQQQQQNEQMGHGNLONTQQQQQSDNEPAPQQQQQ--PTKSYPR--ARK 595
Db 108 QACLOMQOQHHPDPQQQGH--QLQQQQQQQ-----PQQQKPPPMAYPQNIVRP 157
QY 596 SQGGTSGSPSPGQISGSKSRPFAADEVDSNINNAPEQMTTTTTTTRITVTTQTRDG 655
Db 158 GQONADSS-----EVDLDVLNA-----AIRAVITNKPEDG 187

RESULT 25

ABB71899
ID ABB71899 standard; protein; 506 AA.
XX ABB71899;
AC ABB71899;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 42489.
DE Drosophila; developmental biology; cell signalling; insecticide;
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US009231.
PF 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
FA Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-65860/75.
PI N-PSDB; ABL16002.
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.
PT Disclosure; SEQ ID NO 42489; 2lpp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (ABB57737-ABB2072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 506 AA;

Query Match 4.6%; Score 168.5; DB 4; Length 506;
Best Local Similarity 20.6%; Pred. No. 8.6e-05;
Matches 114; Conservative 72; Mismatches 239; Indels 129; Gaps 22;
QY 84 QHLDSSAANQATEKVFQMSFMYRSSAQH--DQKMWREDFAPVYVINSRSQSHG 141
Db 3 QOVNELNISTQAMTSTTT--QEKVGVGVNVLVERQQCEMEFETSSDLVVGQSPSESK 61
QY 142 RTKSGIEKEKHTPMVAPSSHHISIRFQEVNQTGSKQNVCLATCSKPEVDRQVKANARSGF 201
Db 62 KVASDASATVTAPOVQAPAH-----VDSSPQASGLSLLAAYSSDDSDDEKVTQVNGD 114
QY 202 VTSLDVSVTE-----EIDLEKSASS-----HDRVNDYNASLRQESNRRL 240
Db 115 NDVIEVPVTDPAASSTAYRPPVAVSSDDESKSSSSSDSDSDSEGYLTVLKKIDKRI 174
QY 241 YRDGKTKLKDTCAGESHLATENHSGHSGSPEDINDREYSKSRACASLQOI----- 294
Db 175 -----NTVDCDEDEDF-----DEDGATGDRSRRRQPPKVRGEMLLDLPPTHQL 219
QY 295 -----NEASDVSDSDSDSVSISSIDVSPDDVVGILGQKRFW--RARKAIAQORVFAV-- 346
Db 220 EITVPEDECVELGKQVSIQVQLVLSVLPNSMLPDLDTVLFEKGRKVLG---EVDVLG 276
QY 347 QLFE-----LHRLIKVKQLIAASPDLLLDLDEISFGLKV-----SAKSYFVKKL 388
Db 277 QVSDPLYCVRFNSNKKQLIDRGIKIGDVVYCAPKTEHTQFVILSKLMQVRGSDASW----- 331
QY 389 LPSEFLVKPPLPHVYVVGKQDSEKTDQHKMESSAENVVGRISNQGHQOS-----NYMPFA 444
Db 332 ---BHDVEPPARYV-----DHSDDDEEREAREARQKRQRDRDRTNSTDSDVTTSVATT 381
QY 445 NNPPASPAPNGYCFPP-----QPPPSGNHQQLIPVMSPSGLIYKP--H 487
Db 382 ATKASSVAP-----PPRQGRRGQRESFRQQRPSINQHNQ-----NQPDQYNEHPSYN 432
QY 488 PGMATGHYGGYGYHYMPTMVMQYHPGNGFPPPGNGYFPPYGMPTIMNPYCSSQQQQ 547
Db 433 PGWSHSNYQNY--HQAANFNMAQOHPGMPFPVNYGYGMPYAMPMPYHPYPPPPPPFA 490
QY 548 QQQNEQMNQFGHP 561
Db 491 PPPNNQSHQ--GQP 503

Search completed: July 29, 2004, 08:22:21
Job time : 103 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 08:22:26 ; Search time 41 Seconds
(without alignments)

875.123 Million cell updates/sec

Title: US-10-719-885-2

Perfect score: 3669

Sequence:

1 MKRGDEPKLEPFMPRLHV.....AARIFQIERKRYDSSKP 695

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3669	100.0	695	4	US-09-513-057C-2
2	3669	100.0	695	4	US-09-513-057C-35
3	3321	90.5	695	4	US-09-234-827B-2
4	2806	76.5	595	4	US-09-234-827B-4
5	2609.5	71.1	653	4	US-09-513-057C-13
6	773.5	21.1	760	4	US-09-513-057C-27
7	621.5	16.9	389	4	US-09-513-057C-23
8	436.5	11.9	540	4	US-09-513-057C-33
9	397	10.8	248	4	US-09-513-057C-29
10	281.5	7.7	179	4	US-09-513-057C-24
11	231	6.3	49	4	US-09-513-057C-25
12	202.5	5.5	185	4	US-09-513-057C-19
13	189	5.2	116	4	US-09-513-057C-16
14	170.5	4.6	189	4	US-09-513-057C-31
15	162	4.4	816	3	US-08-267-803B-9
16	162	4.4	816	3	US-09-041-886-17
17	158.5	4.3	1061	3	US-08-701-154A-5
18	155.5	4.2	1036	3	US-08-891-640-3
19	154.5	4.2	1402	4	US-09-125-635-12
20	150.5	4.1	528	4	US-09-506-066E-2
21	150.5	4.1	2441	1	US-08-194-468-2
22	150.5	4.1	2441	3	US-08-961-739-2
23	150.5	4.1	2441	4	US-09-514-247A-8
24	150.5	4.1	2441	4	US-09-686-316-2
25	148.5	4.0	1253	4	US-09-489-039A-12097
26	148	4.0	1005	2	US-08-935-450-2
27	148	4.0	1005	4	US-09-338-123-2

1214	4	US-10-164-595-24	Sequence 24, Appl
1214	5	PCT-US93-03027-3	Sequence 3, Appl
1274	4	US-09-095-443-2	Sequence 2, Appl
2442	4	US-09-514-247A-10	Sequence 10, Appl
3969	3	US-08-061-376-5	Sequence 5, Appl
720	4	US-09-508-824-11	Sequence 11, Appl
1312	3	US-09-041-886-19	Sequence 19, Appl
1312	4	US-09-648-281-2	Sequence 2, Appl
1312	4	US-09-707-919A-19	Sequence 19, Appl
1312	4	US-09-083-268-3	Sequence 3, Appl
612	4	US-08-894-454-163	Sequence 163, App
788	2	US-08-918-914-4	Sequence 4, Appl
729	4	US-09-625-188-20	Sequence 20, Appl
1031	4	US-09-914-259-24	Sequence 24, Appl
776	4	US-09-266-225D-10	Sequence 10, Appl
1507	4	US-09-914-259-37	Sequence 37, Appl
2843	1	US-07-741-940-2	Sequence 2, Appl
2843	1	US-08-289-548A-2	Sequence 2, Appl
2843	1	US-08-452-654-2	Sequence 2, Appl
2843	2	US-08-370-235A-2	Sequence 2, Appl
2843	4	US-08-449-731-2	Sequence 2, Appl
703	3	US-08-910-925-4	Sequence 4, Appl
2842	1	US-07-741-940-7	Sequence 7, Appl
2842	1	US-08-289-548A-7	Sequence 7, Appl
2842	1	US-08-452-654-7	Sequence 7, Appl
2842	4	US-08-449-731-7	Sequence 7, Appl
2843	1	US-08-452-655B-2	Sequence 2, Appl
2843	1	US-08-452-655B-7	Sequence 7, Appl
2843	3	US-08-450-582-7	Sequence 2, Appl
2843	3	US-08-450-582-7	Sequence 7, Appl
2973	2	US-08-821-355A-7	Sequence 7, Appl
2973	3	US-09-003-687A-7	Sequence 7, Appl
949	3	US-09-136-605-7	Sequence 7, Appl
1093	4	US-10-164-595-65	Sequence 65, Appl
1093	5	PCT-US94-04496-55	Sequence 55, Appl
796	3	US-08-857-076-40	Sequence 40, Appl
858	3	US-08-857-076-41	Sequence 41, Appl
892	3	US-08-857-076-42	Sequence 42, Appl
1404	4	US-08-801-308-1	Sequence 1, Appl
1850	4	US-09-620-093A-5	Sequence 5, Appl
2074	4	US-09-491-356C-9	Sequence 9, Appl
748	3	US-08-725-459B-27	Sequence 27, Appl
2023	4	US-09-491-356C-8	Sequence 8, Appl
1395	3	US-09-540-245A-15	Sequence 15, Appl
380	2	US-09-026-587-4	Sequence 4, Appl
380	2	US-09-227-420-4	Sequence 4, Appl
380	4	US-09-387-811-4	Sequence 4, Appl
635	4	US-09-417-197-125	Sequence 125, App
753	3	US-08-725-459B-25	Sequence 25, Appl
1184	4	US-09-266-225D-18	Sequence 18, Appl
1185	3	US-09-041-886-23	Sequence 23, Appl
2414	1	US-08-227-536-2	Sequence 2, Appl
2414	5	PCT-US95-04682-2	Sequence 2, Appl
460	4	US-09-056-556-184	Sequence 184, App
460	4	US-09-072-596-179	Sequence 179, App
460	4	US-09-072-967-184	Sequence 184, App
1096	4	US-09-415-946-3	Sequence 3, Appl
2185	4	US-08-891-640-2	Sequence 2, Appl
2245	4	US-09-854-856-36	Sequence 36, Appl
2322	4	US-09-854-856-34	Sequence 34, Appl
2382	4	US-09-854-856-2	Sequence 2, Appl
716	4	US-09-219-983A-20	Sequence 20, Appl
1155	1	US-08-094-948A-29	Sequence 29, Appl
1155	5	PCT-US96-09319-29	Sequence 29, Appl
2157	4	US-09-854-856-52	Sequence 52, Appl
2217	4	US-09-854-856-20	Sequence 20, Appl
2294	4	US-09-854-856-50	Sequence 50, Appl
2354	4	US-09-854-856-18	Sequence 18, Appl
747	3	US-08-973-273-3	Sequence 3, Appl
2004	4	US-09-854-856-58	Sequence 58, Appl
2064	4	US-09-854-856-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-513-057C-2
; Sequence 2, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-513-057C-2

Query Match 100.0%; Score 3669; DB 4; Length 695;
Best Local Similarity 100.0%; Pred. No. 7.4e-308;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRGDDEKILPMFRLHVNDADKGGPRAPRNKMALEYQLSIPQRFGDHGTMSRSN 60
DB 1 MKRGDDEKILPMFRLHVNDADKGGPRAPRNKMALEYQLSIPQRFGDHGTMSRSN 60

QY 61 NTSTLVHPGPSSQPCGVERNLSVQHLDSAAQAOTEKFVSMFMENVRSSAHDQKMW 120
DB 61 NTSTLVHPGPSSQPCGVERNLSVQHLDSAAQAOTEKFVSMFMENVRSSAHDQKMW 120

QY 121 REEDFAVPVYINRSRQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKNVCL 180
DB 121 REEDFAVPVYINRSRQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKNVCL 180

QY 181 ATCSKEPEVRDQVKANARSGGFVLSLDVSVTEEIDLEKSASSHDRVNDYNASLQESNRNL 240
DB 181 ATCSKEPEVRDQVKANARSGGFVLSLDVSVTEEIDLEKSASSHDRVNDYNASLQESNRNL 240

QY 241 YRDGGKTRLKDTNGAESHATENHSQEGHGSPEIDNDREYSKSRACASLQINEEASD 300
DB 241 YRDGGKTRLKDTNGAESHATENHSQEGHGSPEIDNDREYSKSRACASLQINEEASD 300

QY 301 DVSDDSMWDSISSIDVSPDDVVGILGQKRFWRARKAIAHQORVFAVOLFELHRLIKVOKL 360
DB 301 DVSDDSMWDSISSIDVSPDDVVGILGQKRFWRARKAIAHQORVFAVOLFELHRLIKVOKL 360

QY 361 IAASPDLLLDDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVVKQGDSEKTDQHKMES 420
DB 361 IAASPDLLLDDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVVKQGDSEKTDQHKMES 420

QY 421 SAENVVGRISNQGHQOOSNTMPFANNPPASPANGYCFPPQPPSGNHQOQLIPVMSPE 480
DB 421 SAENVVGRISNQGHQOOSNTMPFANNPPASPANGYCFPPQPPSGNHQOQLIPVMSPE 480

QY 481 GLIYKHPGMAHTGHYGGYGYHYMPTPMVMPQVHPGMRPPGNGYFPFPGMPTINPNY 540
DB 481 GLIYKHPGMAHTGHYGGYGYHYMPTPMVMPQVHPGMRPPGNGYFPFPGMPTINPNY 540

QY 541 CSSOQOQOQOQNEQMNQFGHGNLQNTQQOQOQSDNEPAPQOQOQPTKSYPRARKSRQGS 600
DB 541 CSSOQOQOQOQNEQMNQFGHGNLQNTQQOQOQSDNEPAPQOQOQPTKSYPRARKSRQGS 600

QY 601 TGSSPSGQIGSGSKSRPFAAVDESDNINNAPEQMTTTTTTTTTTTTTRTTTQTTTRDGGVTR 660
DB 601 TGSSPSGQIGSGSKSRPFAAVDESDNINNAPEQMTTTTTTTTTTTTTRTTTQTTTRDGGVTR 660

QY 661 VIKVVPHNKLASENAARI FQSIQERKRYDSSKP 695
DB 661 VIKVVPHNKLASENAARI FQSIQERKRYDSSKP 695

DB 661 VIKVVPHNKLASENAARI FQSIQERKRYDSSKP 695

RESULT 2
US-09-513-057C-35
; Sequence 35, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-513-057C-35

Query Match 100.0%; Score 3669; DB 4; Length 695;
Best Local Similarity 100.0%; Pred. No. 7.4e-308;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRGDDEKILPMFRLHVNDADKGGPRAPRNKMALEYQLSIPQRFGDHGTMSRSN 60
DB 1 MKRGDDEKILPMFRLHVNDADKGGPRAPRNKMALEYQLSIPQRFGDHGTMSRSN 60

QY 61 NTSTLVHPGPSSQPCGVERNLSVQHLDSAAQAOTEKFVSMFMENVRSSAHDQKMW 120
DB 61 NTSTLVHPGPSSQPCGVERNLSVQHLDSAAQAOTEKFVSMFMENVRSSAHDQKMW 120

QY 121 REEDFAVPVYINRSRQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKNVCL 180
DB 121 REEDFAVPVYINRSRQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKNVCL 180

QY 181 ATCSKEPEVRDQVKANARSGGFVLSLDVSVTEEIDLEKSASSHDRVNDYNASLQESNRNL 240
DB 181 ATCSKEPEVRDQVKANARSGGFVLSLDVSVTEEIDLEKSASSHDRVNDYNASLQESNRNL 240

QY 241 YRDGGKTRLKDTNGAESHATENHSQEGHGSPEIDNDREYSKSRACASLQINEEASD 300
DB 241 YRDGGKTRLKDTNGAESHATENHSQEGHGSPEIDNDREYSKSRACASLQINEEASD 300

QY 301 DVSDDSMWDSISSIDVSPDDVVGILGQKRFWRARKAIAHQORVFAVOLFELHRLIKVOKL 360
DB 301 DVSDDSMWDSISSIDVSPDDVVGILGQKRFWRARKAIAHQORVFAVOLFELHRLIKVOKL 360

QY 361 IAASPDLLLDDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVVKQGDSEKTDQHKMES 420
DB 361 IAASPDLLLDDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVVKQGDSEKTDQHKMES 420

QY 421 SAENVVGRISNQGHQOOSNTMPFANNPPASPANGYCFPPQPPSGNHQOQLIPVMSPE 480
DB 421 SAENVVGRISNQGHQOOSNTMPFANNPPASPANGYCFPPQPPSGNHQOQLIPVMSPE 480

QY 481 GLIYKHPGMAHTGHYGGYGYHYMPTPMVMPQVHPGMRPPGNGYFPFPGMPTINPNY 540
DB 481 GLIYKHPGMAHTGHYGGYGYHYMPTPMVMPQVHPGMRPPGNGYFPFPGMPTINPNY 540

QY 541 CSSOQOQOQOQNEQMNQFGHGNLQNTQQOQOQSDNEPAPQOQOQPTKSYPRARKSRQGS 600
DB 541 CSSOQOQOQOQNEQMNQFGHGNLQNTQQOQOQSDNEPAPQOQOQPTKSYPRARKSRQGS 600

QY 601 TGSSPSGQIGSGSKSRPFAAVDESDNINNAPEQMTTTTTTTTTTTTTRTTTQTTTRDGGVTR 660
DB 601 TGSSPSGQIGSGSKSRPFAAVDESDNINNAPEQMTTTTTTTTTTTTTRTTTQTTTRDGGVTR 660

QY 661 VIKVVPHNKLASENAARI FQSIQERKRYDSSKP 695
DB 661 VIKVVPHNKLASENAARI FQSIQERKRYDSSKP 695

```
RESULT 3
US-09-234-827B-2
; Sequence 2, Application US/09234827B
; Patent No. 6448471
; GENERAL INFORMATION:
; APPLICANT: Puzio, Piotr S.
; APPLICANT: Grundler, Florian M.W.
; TITLE OF INVENTION: Nematode feeding structure specific gene and its
; FILE OF INVENTION: application to produce nematode resistant plants
; FILE REFERENCE: U-012084-2
; CURRENT APPLICATION NUMBER: US/09/234,827B
; CURRENT FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: US 60/072,142
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-234-827B-2

Query Match      90.5%; Score 3321; DB 4; Length 695;
Best Local Similarity 91.0%; Pred. No. 7.7e-278;
Matches 648; Conservative 4; Mismatches 26; Indels 34; Gaps 3;

QY 1 M K R G D E E K I L P M F P R L H V N D A D K G P R A P P R N K M A L Y E Q L S I P S O R F G D H G T M N S R N 60
Db 1 M K R G D E E K I L P M F P R L H V N D A D K G P R A P P R N K M A L Y E Q L S I P S O R F G D H G T M N S R N 60
QY 61 N T S T L V H P G S Q P C G V E R N L S V Q H L D S A A N Q A T E K F V S Q M S F M E N V R S S A Q H D Q R K M V 120
Db 61 N T S T L V H P G S Q P C G V E R N L S V Q H L D S A A N Q A T E K F V S Q M S F M E N V R S S A Q H D Q R K M V 120
QY 121 R E E D F A V P V I N S R S Q S H G R T K S I E K E K T P M A P S H S I R F Q V N O T G S K O N V C L 180
Db 121 R E E D F A V P V I N S R S Q S H G R T K S I E K E K T P M A P S H S I R F Q V N O T G S K O N V C L 180
QY 181 A T C S K P E V R D Q V K A N A R S G G F V I S L D V S T E I D L E K S A S S H D R V N D Y N A S I R Q S R N R L 240
Db 181 A T C S K P E V R D Q V K A N R S G G F V I S L D V S T E I D L E K S A S S H D R V N D Y N A S I R Q S R N R L 240
QY 241 Y R D G K T R L K D T N G A E S H L A T E N H S Q E G H G S P E D I D N D R E Y S K S R A C A S L O Q I N E E A S D 300
Db 241 Y R D G K T R L K D T N G A E S H L A T E N H S Q E G H G S P E D I D N D R E Y S K S R A C A S L O Q I N E E A S D 300
QY 301 D V S D S M V D S I S S I D V S P D V V G I L G Q K R F W A R K A I A N Q O R V A V O L F E L H R L I K V Q K L 360
Db 301 D V S D S M V D S I S S I D V S P D V V G I L G Q K R F W A R K A I A N Q O R V A V O L F E L H R L I K V Q K L 360
QY 361 I A A S P D L L D E I S F L G K V S A K S Y P V K K L P S E F L V K P P L P H V V V K O R G D S E K T D O H K M S 420
Db 361 I A A S P D L L D E I S F L G K V S A K S Y P V K K L P S E F L V K P P L P H V V V K O R G D S E K T D O H K M S 420
QY 421 S A E N V V G R L S N G H H Q H P T T C L F ----- Q Q T T H R L H R L Q M D I A F L L S L 463
Db 421 S A E N V V G R L S N G H H Q H P T T C L F ----- Q Q T T H R L H R L Q M D I A F L L S L 463
QY 470 -- Q W L I P V M S P S E G L I ----- Y K P H P G M A H T H Y G G Y G H Y M T P M V M P O Y H P G M G F P P P G 523
Db 470 -- Q W L I P V M S P S E G L I ----- Y K P H P G M A H T H Y G G Y G H Y M T P M V M P O Y H P G M G F P P P G 523
QY 524 N G Y F P P Y G M P T I M N P Y C S S Q Q Q Q Q Q P N E Q M N Q F G H P G N L O N T Q Q Q Q R S D N E P A P Q Q 583
Db 524 N G Y F P P Y G M P T I M N P Y C S S Q Q Q Q Q Q P N E Q M N Q F G H P G N L O N T Q Q Q Q R S D N E P A P Q Q 583
QY 584 Q O P T K S Y P R A K S R O G S T G S S P G Q I S G S K S F R P F A A V D E D S N I N N A P E Q T M T T T T T 643
Db 584 Q O P T K S Y P R A K S R O G S T G S S P G Q I S G S K S F R P F A A V D E D S N I N N A P E Q T M T T T T T 643
QY 644 T R T T V T Q T T R D G G V T R V I K V V P H N A K L A S E N A A R I F Q S I Q E E R K R Y D S S K P 695
Db 644 T R T T V T Q T T R D G G V T R V I K V V P H N A K L A S E N A A R I F Q S I Q E E R K R Y D S S K P 695

RESULT 4
US-09-234-827B-4
; Sequence 4, Application US/09234827B
; Patent No. 6448471
; GENERAL INFORMATION:
; APPLICANT: Puzio, Piotr S.
; APPLICANT: Grundler, Florian M.W.
; TITLE OF INVENTION: Nematode feeding structure specific gene and its
; FILE OF INVENTION: application to produce nematode resistant plants
; FILE REFERENCE: U-012084-2
; CURRENT APPLICATION NUMBER: US/09/234,827B
; CURRENT FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: US 60/072,142
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-234-827B-4

Query Match      76.5%; Score 2806; DB 4; Length 595;
Best Local Similarity 89.9%; Pred. No. 1.6e-233;
Matches 550; Conservative 4; Mismatches 24; Indels 34; Gaps 3;

QY 101 Q M S F M E N V R S S A Q H D Q R K M V R E E D F A V P V I N S R R S Q S H G R T K S I E K E K T P M V A P S S 160
Db 1 Q M S F M E N V R S S A Q H D Q R K M V R E E D F A V P V I N S R R S Q S H G R T K S I E K E K T P M V A P S S 160
QY 161 H H S I R F Q E V N O T G S K O N V C L A T C S K P E V R D Q V K A N A R S G G F V I S L D V S T E I D L E K S A S 220
Db 61 H H S I R F Q E V N O T G S K O N V C L A T C S K P E V R D Q V K A N R R S G G F V I S L D V S T E I D L E K S A S 120
QY 221 S H D R V N D Y N A S I R Q S R N R L Y R D G K T R L K D T N G A E S H L A T E N H S Q E G H G S P E D I D N D R 280
Db 121 S H D R V N D Y N A S I R Q S R N R L Y R D G K T R L K D T N G A E S H L A T E N H S Q E G H G S P E D I D N D R 180
QY 281 E Y S K S R A C A S L O Q I N E E A S D D V S D S M V D S I S S I D V S P D V V G I L G Q K R F W A R K A I A N Q 340
Db 181 E Y S K S R A C A S L O Q I N E E A S D D V S D S M V D S I S S I D V S P D V V G I L G Q K R F W A R K A I A N Q 240
QY 341 Q R V F A V O L F E L H R L I K V Q K L I A A S P D L L D E I S F L G K V S A K S Y P V K K L P S E F L V K P P L P 400
Db 241 Q R V F A V O L F E L H R L I K V Q K L I A A S P D L L D E I S F L G K V S A K S Y P V K K L P S E F L V K P P L P 300
QY 401 H V V V K O R G D S E K T D O H K M E S S A E N V V G R L S N G H H Q Q S N Y M P F A N N P P A S P A P N G Y C F P P 460
Db 301 H V V V K O R G D S E K T D O H K M E S S A E N V V G R L S N G H H Q H P T T C L F ----- 343
QY 461 Q P P S G N H Q ----- Q W L I P V M S P S E G L I ----- Y K P H P G M A H T H Y G G Y G H Y 503
Db 344 Q Q T T H R L H R L Q M D I A F L L S L L Q E I I S N G D P C N V S L G R T D I Q P H P G M A H T H Y G G Y G H Y 403
QY 504 M P T P M V M P O Y H P G M G F P P P G N G Y F P P Y G M P T I M N P Y C S S Q Q Q Q Q Q Q N E Q M N Q F G H P G N 563
Db 404 M P T P M V M P O Y H P G M G F P P P G N G Y F P P Y G M P T I M N P Y C S S Q Q Q Q Q Q Q N E Q M N Q F G H P G N 463
QY 564 L Q N T Q Q Q Q R S D N E P A P Q Q O P T K S Y P R A K S R O G S T G S S P G Q I S G S K S F R P F A A V 623
Db 464 L Q N T Q Q Q Q R S D N E P A P Q Q O P T K S Y P R A K S R O G S T G S S P G Q I S G S K S F R P F A A V 523
QY 624 D E D S N I N N A P E Q T M T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T 683
Db 524 D E D S N I N N A P E Q T M T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T 583
QY 684 Q E E R K R Y D S S K P 695
Db 584 Q E E R K R Y D S S K P 595
```

RESULT 5

US-09-513-057C-13
; Sequence 13, Application US/09513057C
; Patent No. 6433251

; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.

; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C

; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin version 3.1

; SEQ ID NO 13
; TYPE: PRT
; LENGTH: 653

; ORGANISM: Cardamine oligosperma
; ORGANISM: Cardamine oligosperma

US-09-513-057C-13
Query Match 71.1%; Score 2609.5; DB 4; Length 653;
Best Local Similarity 76.7%; Pred. No. 1.7e-216;
Matches 517; Conservative 47; Mismatches 67; Indels 43; Gaps 12;

QY 1 MKRGKDEKLEPMFPRHLVNDADKGGPRAPPENKMAVLEQLSIPSORFGDHG--TWNSR 58
DB 1 MKRGKDEKLEPMFPRHLVNDADKGGPRAPPENKMAVLEQLSIPSORFGDHGNSLSR 60
QY 59 SNNSTSLVHPGPS-OPCGVERNLVQHLDSSAANOATEKFVSQMFENVRSSAQHDQR 117
DB 61 SNNSTSLVHPGPSQSCGVERNLVQHLDSSAAVHTENFVSQMPFEMNMRSLAKHDQR 120
QY 118 KMYREEDFAVPVYINRRSQSGHRTKSGTEKHKFTPMVAPSSHHSIRFQVNTGSKQN 177
DB 121 KTVREEDFAVPVFNRRSQSGHRTKSGTGVIEKHT----- 156
QY 178 VCLATCSKEPVROVKANRSGGFVSLDVSVEEIDLEKSASHDRVNDYNASLROESR 237
DB 157 -TLATCSK-LVRDKVKNNAKSGGFI---DLSTEEVDLEKSASYDRVNDYCNLSLROESR 211
QY 238 NLYRDGGKRLKDTONGAESHLATENHSOEGHSGSPEDINDREYSKSRACASLQINEE 297
DB 212 NKLYRDGGGAHMDTANRVESHLVTESHSEHGSPDDDDNGHEYCSRGGVSLQINEE 271
QY 298 ASDVDSDSMVDSTSSIDVSPDDVVGILGQKRFWRARAKATANOQRFVAVQLFELHRLIKV 357
DB 272 ASDVDSDSMVDSTSSIDVSPDDVVGILGQKRFWRARAKATANOQRFVAVQLFELHRLIKV 331
QY 358 OKLTAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVY--KORGDSKTDQ 415
DB 332 ORLTAASPDVLDDEINYLKGVSAKSYPVKKLVSEFIVKPLPQVNVNKHRSKTDQ 391
QY 416 HKMESSAENVVGRLSNOGH---HQSNMPPFANNPPASPAANGCYFPFPPPSNGHQWL 472
DB 392 HKMCSAENVVGRLSNOGHNNHQPNSNMPSPSPASPAVNGCYFPFPPPSGN-QQWL 450
QY 473 IPVMSPEGGLYKHPHGMVHTCHGGYGYHYMPTPMVMPQYHPCGPPPGNGYFPPYGM 532
DB 451 IPVMSPEGGLYKHPHGMVHTCHGGYGYHYMPTPMVMPQYHPCGPPPGNGYFPPYGV 510
QY 533 MPTTMNPPCSSQQQQQ--QPNEMQNOFGHPCNLQNTQQQQQSDNEPAPQQQQQPTKSYP 591
DB 511 IPAMNPPYGPQQQQQQQQAQNEQNTQNGYSGNLQNNTHQES-SVNEAAPQPEL-TKSYP 568
QY 592 RARKSRQSGTSGSPGQISGSKSFRFAAVDSGSNN--APEQMTTTTTTTTTTTT 649
DB 569 RARKSRQSVTASSASGREGISGTSFRPFSFSAVDENNDNNDADQMNTTTTTTTTTT 628
QY 650 QTRDGGGVTRVIK 663
DB 629 QTRDGGGVTRVIK 642

RESULT 6

US-09-513-057C-27
; Sequence 27, Application US/09513057C
; Patent No. 6433251

; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.

; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C

; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin version 3.1

; SEQ ID NO 27
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Oryza sativa

US-09-513-057C-27
Query Match 21.1%; Score 773.5; DB 4; Length 760;
Best Local Similarity 32.1%; Pred. No. 4.9e-58;
Matches 261; Conservative 105; Mismatches 257; Indels 191; Gaps 36;

QY 4 GKDEE-KILEMPFPRHLVNDADK-GGPRAPPENKMAVLEQLSIPSORF---GDHGTMS 57
DB 12 GKDEE-KILEMPFPRHLVNDADK-GGPRAPPENKMAVLEQLSIPSORF---GDHGTMS 71
QY 58 RSNTSTLVHPGPSOPCGVERNL-----SVQHLDSSAANOATEKFVSQMF 104
DB 72 PHSTSAASQSQSQVYGRDSSLFQPFNVPSNRPGHSTKINSKINKISGRKELGM 131
QY 105 MEN-----VRSAQHDQR-----KMYREEDFAVPVYINRRSQSGHRTK 144
DB 132 LSSQTKGMDIYASRSATAPORRAENTIKSSGKRLADDDFVMPVFNRRSQSGHRTK 191
QY 145 SGLEKEKHTPMVAPSSHHSIRFQVNTGSKQN-CLATCSKEPVROVK-ANARSGGFV 202
DB 192 AGVQ-DOSTPLVAANPHKS-----PSTVSKSTCYNTVSKLERLHSDVKSRTPLKD 244
QY 203 ISLDVSVT-EEIDLEKSASHDRVNDYNASLROESR-RLYRDGGKTR-LKQTD---NGA 256
DB 245 KEMEAQTSKNVEVEKSSSFHASKDMF-----ESRHAQVYPMKDKTGIIINDSEPHG 298
QY 257 ESHLATENHS-----OEGHSGPEDIINDREY-----SKSRACASLQOIN 295
DB 299 SCHQATSRNGSGMKFQNPMPMRNEISSNPSENDRHYNLPQGGIETGTRKRLLEQHD 358
QY 296 EASDDVS-----DDSMVDSTSSIDVSPDDVVGILGQKRFWRARAKAIANQQ 341
DB 359 AEKSDVSRLLLEQHDENIDDDVSDSVCEITGWEISPDKIVGALGTGKFWKARRAIMNQ 418
QY 342 RVFAVQLFELHRLIKVOKLTAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPPLPH 401
DB 419 RVFAVQLFELHRLIKVOKLTAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPPLPH 475
QY 402 VVVKQSGDSEKTOHKKVSSAENV-----VGRLSNOGHQCSNMPFAN---NPPA 449
DB 476 ATI-----DDVEPSLQOPEVSKENTEDSPPHDTGLSGGQDQATNGVSKSNRRATPVA 531
QY 450 SP-APNGCYCPPQPPSGNHOQLIPVMSPEGGLYKHPHG-----M 490
DB 532 SDNKQNWGVLQDPP-----QNOVLVPMVSLGLELVKPYSGPCPPAGSILLAFVANCPTL 587
QY 491 AHTGHYGGYGYHYMPTPMVMPQYHPCGFPB--PGNGYFPPYGMPTMNPYCSQQQQQ 548
DB 588 SLPTAGDFMNSAYGVMPHQPQHMGAFFGPPSMEN--YFPFFSI--PVMNP----- 635
QY 549 QOPNEQNOFGHPCNLQ---NTOQQQORSNNEPAPQQQQQPTKSYP---RARKSR----- 597
DB 636 TAPAPVVEQGRHPSNPQPYGNFEQQSWISCN-----MSHPSGIWRFHARSDEA 684
QY 598 QGSTGSGSPGQISGSKSFRFAAVDSGSNNAPQMTTTTTTTTTTTTTRDGGG 657
DB 685 QASSASSFPDRFQCSGS---GPVSAPFTVSAQNNQOP-----SYSRD--N 726

```
QY 658 VTRVIVKVPNNHNAKLASANAARIFQSIQEEKRYD 691
Db 727 QTNVIVKVPNNHNGRTASASAARIFRSIQMERQDD 760

RESULT 7
US-09-513-057C-23
; Sequence 23, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-513-057C-23

Query Match 16.9%; Score 621.5; DB 4; Length 389;
Best Local Similarity 39.0%; Pred. No. 2.4e-45;
Matches 164; Conservative 58; Mismatches 132; Indels 67; Gaps 17;

QY 297 EASDDSDSDSDVSSIDVSDVGVILGQRFWRKAKAIANQORVFAVOLFELHRLIK 356
Db 2 DRGDDLSRVSISGTDISDDIVGIIGLRFRKARRAIVQORVFAIVFELHRLIK 61
QY 357 VOKLIAASPDLLDLDEISFLGKVSASYPVKLLPSEELVKPLPHVHVVKQSGDSEKTDQH 416
Db 62 VQRLIAGSPNSLSLEDPAYLGK-PLKSSSIKR-LPLDCIVRE--SQSVLKRKHDSK-PHF 116
QY 417 KMESSAENVVGRLS--NOGHHQOQNYMPFANNPPASAPNG-----YCFPPQPPPSGNH 468
Db 117 RMEHTAESNVGRASISLVQNGSLSKHPFGTPLPTPTVNDNSAGPWCFF--QQPSGH- 172
QY 469 QWLIVPMSPSGLIYKHPGMAHTG-----HYGGYGHYMTPTVMVQYHPG 516
Db 173 -QWLIVPMSPSGLIYKPSGFGTSPICGSGPSSSPTMGFFAFTYGVPAFNPHYQ-G 230
QY 517 MGF--PPPGNYFPYPMPTIMNPYCSSQQQQQQQNEQNNQFHPGNLQN-----T 567
Db 231 MGVFPAPPTGCHYFRQYGM--FAMNPPISS-----TASEESNQYTMPG-LQHFSGVVD 281
QY 568 QQQQORSNDNEPAPQQQQQPTKSYPRARKSRQGST-GSSPSGQGISGSKSFRPFAAVDED 626
Db 282 DVQHSTSGLSNVLNQKENVPDVRYQSTKDNVEVQASSASSPIETAGRNMLSLFPTSPVT 341
QY 627 SNINNAPEQTMTTTTTTTTTTTTQTRDGGGVTRIVKVVPHNAKLASANAARIFOSIQE 686
Db 342 DNRDGSQACVDPNP-----ARVIVKVVPHNARSATESVARIFOSIQE 384
QY 687 R 687
Db 385 R 385

RESULT 8
US-09-513-057C-33
; Sequence 33, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33

QY 658 VTRVIVKVPNNHNAKLASANAARIFQSIQEEKRYD 691
Db 727 QTNVIVKVPNNHNGRTASASAARIFRSIQMERQDD 760

Query Match 11.9%; Score 436.5; DB 4; Length 540;
Best Local Similarity 23.6%; Pred. No. 3.5e-29;
Matches 177; Conservative 98; Mismatches 195; Indels 279; Gaps 32;

QY 5 KDEEK--LLEPMFPRLLHVNDADKGGPRAPPNNKMAIYEQLS-----TPSQRFQDGHGTWN 56
Db 5 KDEAKRIITIPPLFPFVHVNDTGRGG-----LSQOFDQKTSWLSVSSKRPNLPSPTN 54
QY 57 SRSNNTSTL--VHPGPSSQPC--GVERNLSVQHLDSSAANQATEKFVS---QMSFMEN 107
Db 55 NISDSLSLTFSLSLPPPPNNARLIDGPKNQSPINTKFEGLKLNKNGINYTFPKSSSVTN 114
QY 108 VRSSA--QHDQKVMREBEDFAVPVYIINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIR 165
Db 115 TKPSSIKQNEVYKLNLTSLDSIKSPIVIHSE-----IDPQANTDL-----SLQ 156
QY 166 FQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVLSLDVSVTEERIDLEKSASHORV 225
Db 157 F-----CTSGSKP-----GGEAVVSGKILLSE-----RL 181
QY 226 NDYNASLRQESNRRLYRDGGKTRLKDTONGAESHLATENHSGHGSPEIDINDREYSKS 285
Db 182 EDEN-----QNGSPNVMKTSQYR-----NFAEFNNEQ-KKP 213
QY 286 RACASLQINBEASDDSDSDSDVSDVSSIDVSDVGVILGQRFWRKAKAIANQORVFA 345
Db 214 KTLPRREQV-----ASNCSAIESLSGTSASSYDIARVIGEKRFWKRTYMINQOKIFA 266
QY 346 VOLFELHRLIKVOKLIAASPDLLDLDEISFLGKVSASYPVKLLPSEELVKPLPHVHVVK 405
Db 267 GOVFELHRLIMVQKMAKSPNLFLES-----K 293
QY 406 QRGDSEKT--DQHKMESAENVVGRLSNQGHQOQSNYMPFANNPPASAPNGY----- 456
Db 294 LNGVKHGTWRSSHQLAMAAKV--RKPTENHK-----PVPEYEPHMKPK 337
QY 457 -----CFPPQ--PPPSGNHQWLIPVMSPSGLIYKHPGMAHTGHHYGGYGH 502
Db 338 LPLPSISKELVTPIWPPQQLLPPGN--QWLVPVITSDGLVYKPPFG----- 382
QY 503 YMTPT--WVMPQY-----HPGMGEP--PPGNGYFPYPMPTIMNPYCSSQQQQQQ 549
Db 383 --PCPPSSSAFMPVYVYQDSLETFRFPVSSPFSHSYFPPPNARTTV----- 427
QY 550 QPNEQMNQFHPGNLQNTQQQQORSNDNEPAPQQQQQP---TKSYPRARKSRQSGTSSPS 606
Db 428 ---DQTNPPG-----QFQWNTSSHTQAIPELSKKSQESNDSIDHGSTASSP- 473
QY 607 GPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTTQTRTQTT---RDGGGVTRVI 662
Db 474 -----PEKKHLEVLPLEPTPTHTQTDYKQKQPMRLAI 507
QY 663 KVVPHNAKLASANAARIFOSIQEERKYD 691
Db 508 KAVPHNSTASASAARIFRQIQRERDSD 536

RESULT 9
US-09-513-057C-29
; Sequence 29, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
```



```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (88)..(88)
; NAME/KEY: unsure
; LOCATION: (98)..(98)
; OTHER INFORMATION:
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
; NAME/KEY: unsure
; LOCATION: (113)..(113)
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
; NAME/KEY: unsure
; LOCATION: (114)..(114)
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
US-09-513-057C-29

Query Match 10.8%; Score 397; DB 4; Length 248;
Best Local Similarity 29.5%; Pred. No. 2.9e-26;
Matches 121; Conservative 34; Mismatches 75; Indels 180; Gaps 13;

QY 297 EASDDVSDMSVDSISSIDVSDVVGILQGRFWRKRAKAIANQQRVFAVQVFLHRLIK 356
DB 3 EQNDLSDSS-VESLPGMEISPDVVSALGPKHFKARAIYNQQRVFAVQVFLHRLIK 61
QY 357 VQKLIASPDLLDETSFLGKVSAKSPYVKLLPSFLVKPLPHVHVVKQRGSEKTDQH 416
DB 62 VQKLIASPHVLTEGDPCLGKSLAVSKK----- 89
QY 417 KMESSAENVVGRLSNQGHQOOSNVMFPANNPPASPAPNGYCFPPQPPSGNHQWLIPW 476
DB 90 -----RUS----- 99
QY 477 SPSEGLIYKPHPCMAHTGHYGYGYHYMPTPMVMPOYHP-----GMGFPFPGNG----- 525
DB 100 SPEGLVYKPYKXPSG-----GSLAPPFPAS--YPTSSSSTAGGDFMSSACCARLM 151
QY 526 -----YTPPYGMPTIMNYPYSSQOQQOQNEQWQFHPGNLQNTQQOQQRSDNEPAPQ 581
DB 152 SAPVYFPFS-MPAV-----SGSAVEQVSHVA-----ASQHRNCSSEAVLA 192
QY 582 QOQOPTKSYBRANKSRQSGTSGSPGQISGSKSPRFAAVIDSDSNINNAPEQTMTT 641
DB 193 SRDSEV-----QGSSASFSASE----- 210
QY 642 TTTTTRVTTQTRDGGVTRVYKVVPHNAKLASNAARIFQSIQERKRYD 691
DB 211 -----TAAQ-----PRVIRVVPHARTASESAAIRFASIOMERKQND 247

RESULT 10
US-09-513-057C-24
; Sequence 24, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Lycopodium obscurum
US-09-513-057C-24

Query Match 7.7%; Score 281.5; DB 4; Length 179;
Best Local Similarity 41.1%; Pred. No. 1.7e-16;
Matches 79; Conservative 25; Mismatches 61; Indels 27; Gaps 9;

QY 1 MKRGKDEKILEPWFRLHVNDADKGGPRAPPNNKMALEYQLSIPQRFCDHGTMMNSRN 60
DB 1 MKRGTEGEEKVGMFPFLNVNDTEKGGPRAPPNNKMALEYQLSIPQRY-NPGDLPHNSS 59
QY 61 NTSTLVHPGPSSQPCGVERNLSVQHLDSAAANOATEKFVSQMSF-----MENVTSSAQHDQ 116
DB 60 NSANLVLPHP-SQENHERGVLEFR-QLPALRHPVEKPYGRSSGNTPLREVSKRQ--- 114
QY 117 RKMVREEDFAVPVYINSR-----RSQSHGRTKSGIEKE-----KHTPMVAPSSSHHSIRFOE 168
DB 115 -----TEXEDFRVPTFDSKERAVNTEDYSKGTSDIDKRDSTLTKTDQL-----SHVTPRENL 167
QY 169 VNQTGS--KQNV 178
DB 168 VNTPGESHKINI 179

RESULT 11
US-09-513-057C-25
; Sequence 25, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Lycopodium obscurum
US-09-513-057C-25

Query Match 6.3%; Score 231; DB 4; Length 49;
Best Local Similarity 85.7%; Pred. No. 5.8e-13;
Matches 42; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKRGKDEKILEPWFRLHVNDADKGGPRAPPNNKMALEYQLSIPQRF 49
DB 1 MKRGTEGEEKLMGPFRLHVNDTEKGGPKAPPRNNKMALEYQLSIPQRF 49

RESULT 12
US-09-513-057C-19
; Sequence 19, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Glycine max
US-09-513-057C-19

Query Match 5.5%; Score 202.5; DB 4; Length 185;
Best Local Similarity 33.3%; Pred. No. 1.1e-09;
Matches 69; Conservative 20; Mismatches 69; Indels 49; Gaps 9;

QY 503 YMPETWMPQVHPGNG-----FPPPCNGYFPYPPYGMPTIMNYPYSSQOQQOQNEQWQF 558
DB 10 FNNPAYQFPASHPVPVGVSPFVPPASHTYFAFGM--PVMN-----QATSGSAVEQVNVQF 61
QY 559 GHFGNL-QN-----TQQQQQRSDNEPAPQOQQOQPTKSYPRANKSR--QGSGTSSP 605

```


Qy	575	DNEPAPQQQQP	PTKSYPR	ARKSRQ--	GSTGSSPG	PGQISG	SKSPR	FAA	VDSD	SNINNA	632	
Db	237	PGSPPPAQNQ---	YVHSSSP	QNTGRT	ASPPALP-					VHLH	272	
Qy	633	PQQTMT--	TTTTTTT	TVTQT	TTRGGV	TRVIVK	VP	HN	AKL	ASNA	ARI	684
Db	273	PQQTMLP	HTLT	LGPSQ	VMQYAD	SGS---	HFV	PRE	ATK	KAF--	SSR	326
Qy	685	---	EERKRY	690								
Db	327	GEWKS	RRY	335								

RESULT 17
US-08-701-154A-5
; Sequence 5, Application US/08701154A
; Patent No. 6380373
; GENERAL INFORMATION:
; APPLICANT: O'Malley, Bert W.
; APPLICANT: Tsai, Ming-Jer
; APPLICANT: Tsai, Sophia Y.
; APPLICANT: Onate, Sergio A.
; TITLE OF INVENTION: STEROID RECEPTOR COACTIVATOR
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08701,154A
; FILING DATE: August 21, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/003,784
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/243
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1061 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-701-154A-5

```

; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1061 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-701-154A-5

Query Match          4.3%; Score 158.5; DB 4; Length 1061;
Best Local Similarity 20.1%; Pred. NO. 8.9e-05;
Matches 156; Conservative      89; Mismatches 240; Indels 291; Gaps 37

QY    18 LHVNDADKGGPRAPERNKMALVEQLSIPSRFGDH-----GTMNSRSNNTS   63
      ||| ||||| : : : : : : : : : : : : : : : : : : : : : :
Db     233 LHNNRLSDGDSKYQTSKHLVQLLTITTAEQQLRHADIDTCKDVLSTCGTGSNSASANS 292
      ||| ||||| : : : : : : : : : : : : : : : : : : : : : :
QY     64 TLVHPG---PSSQPCGVERNLVSQHLDSSAAQAQATEKFVQSFMENVRSSAQHQDKMV 120
      ||| ||||| : : : : : : : : : : : : : : : : : : : : : :

```

```
Db 293 -----GSCSPSS-----HSSLTARHKILHRLQEGSPSDDITLISVEPDKDSA 335
QY 121 REEDFAVPVYNS-----RRSQSHGRKTSKIEKEKHTPMVAPSGSHHSIRFQEV 169
Db 336 STSVSVTGQVQGNSSIKLELDASKKESKDHLLRYLLDKDKDLRSTP-----NLSLDDV 391
QY 170 NOTGSKQ-----NVLATCSKPEVRDQVKANARSGGFVISLDSVSTEID-----LE 216
Db 392 KVKVEKQMDPCNTPNTPTKATPE-----EIKLEAQS-QFTADLD-----QFDQLLPTLE 442
QY 217 KSAS-----SHDRVND-----YNASLRQESR-NRLYRDGKGKTR 248
Db 443 KAQLPGLCETDRMDGAVTSVTTKSEILPASLOSATARPPTSRLNRL-----PELELAIDN 498
QY 255 GAESHLATENHSGHGSPEIDNDREYSKSRACASLQINEASDDVSDSDMSVDSI--- 311
Db 499 -----QFGQPGTGDIPTWNN-----TVTAINQSKSEDOCISSQLDELLCP 539
QY 312 -SSIDVSPDDVVGILGKRFWRARKAIAANOQRFV-----AVQLFELHRLIKVKLI--- 361
Db 540 PTTVEGRNDE-----KALLEQLVSLSGKDETELAELDRALGIDKLVQGGG 585
QY 362 -----AASPOLLLDE-----ISFLGKVSASAK-----SYPVK- 386
Db 586 LDVLSERFPQQATPPLIMEERNLYSQYSPFPPTANLPSPPQGMVVRQKPSLGTMPVQV 645
QY 387 -----KLLPSEFLVKPPLP-----HVVVKQR--GDSEKTDQHKM-----ESSAENV 425
Db 646 TPRGAFSPGMGMQPRQTLNRPAPNQLRLQQRLOGQQLHQNRAILNQFAATAP 705
QY 426 VGLSLNQGHQSQSNYP-----FANN-P 447
Db 706 VGINMRSGMQQQ-TPOPPNLAQMLAQORELYSQQRQRLIQOQRAMLRMQOSFGNNLP 765
QY 448 PAS-----PAENGVCFFPQ-----BPPSGNHQQLIPVMSPSGLIYKPH 487
Db 766 PSSGLPVQTNPRLPQGAFOQFPYPNYGNTNPTGPAS-----TSPFSLAANPE 815
QY 488 PGMAH-----TGHYGGYGYHYMPTPMVMPQY-HPGMGFPFPPGNGYFPP-----YGM 532
Db 816 ASLANRNSMVSRGTMGNIGQFGTGINPQMQQNVFOYFGAGVMPQGEANFAPSLSGSSM 875
QY 533 MPTIMNPPYSSQQQQ-----QQQPMQNMQFQHGPNLQNTQQQQQORSNEPAPQ 582
Db 876 VMPPIPPQSSLLQOTPPASGYQSPDMKAWQQAIGNNNNVFSQAVQ---NQPTPAQ 928
```

RESULT 18

```
US-08-891-640-3
; Sequence 3, Application US/08891640
; Patent No. 6268173
; GENERAL INFORMATION:
; APPLICANT: Chambon, Pierre
; APPLICANT: Gronemeyer, Hinrich
; APPLICANT: Voegel, Johannes
; APPLICANT: Lutz, Yves
; TITLE OF INVENTION: Transcriptional Intermediary Factor-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,640
; FILING DATE: Herewith
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/021,247
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1036 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-08-891-640-3
```

Query Match 4.2%; Score 155.5; DB 3; Length 1036;
Best Local Similarity 19.9%; Pred. No. 0.00016;
Matches 156; Conservative 89; Mismatches 240; Indels 297; Gaps 37;

```
QY 18 LHVNDADKGGPRAPPRNKALYEQLSIPSORFGDH-----GTWMSRNNNTS 63
Db 233 LHNDRSLDGDSKYSQTSKHLVQLLTTRAEQQLRHADIDTCKDVLCTGTSNSASANS 292
QY 64 TLVHFG-----PSSQPCGVERNLSVQHLDSSAANOATEKFVQSMFNMENVRSSAQHDQKMW 120
Db 293 -----GSCSPSS-----HSSLTARHKILHRLQEGSPSDDITLISVEPDKDSA 335
QY 121 REEDFAVPVYNS-----RRSQSHGRKTSKIEKEKHTPMVAPSGSHHSIRFQEV 169
Db 336 STSVSVTGQVQGNSSIKLELDASKKESKDHLLRYLLDKDKDLRSTP-----NLSLDDV 391
QY 170 NOTGSKQ-----NVLATCSKPEVRDQVKANARSGGFVISLDSVSTEID-----LE 216
Db 392 KVKVEKQMDPCNTPNTPTKATPE-----EIKLEAQS-QFTADLD-----QFDQLLPTLE 442
QY 217 KSAS-----SHDRVND-----YNASLRQESR-NRLYRDGKGKTR 248
Db 443 KAQLPGLCETDRMDGAVTSVTTKSEILPASLOSATARPPTSRLNRL-----PELE 498
QY 249 LKDTNGAESHLATENHSGHGSPEIDNDREYSKSRACASLQINEASDDVSDSDSV 308
Db 499 LEAIDN-----QFGQPGTGDIPTWNN-----TVTAINQSKSEDOCISSQL 539
QY 309 DSI-----SSIDVSPDDVVGILGKRFWRARKAIAANOQRFV-----AVQLFELHRLIKVK 359
Db 540 DELLCPPTVEGRNDE-----KALLEQLVSLSGKDETELAELDRALGIDK 585
QY 360 LI-----AASPOLLLDE-----ISFLGKVSASAK----- 381
Db 586 LVQGGGLDVLSERFPQQATPPLIMEERNLYSQYSPFPPTANLPSPPQGMVVRQKPSLG 645
QY 382 SYPVK-----KLLPSEFLVKPPLP-----HVVVKQR--GDSEKTDQHKM-----E 419
Db 646 TMPVQVTPPRGAFSPGMGMQPRQTLNRPAPNQLRLQQRLOGQQLHQNRAILNQ 705
QY 420 SSAENVVGLSLNQGHQSQSNYP-----FANN-P 447
Db 706 FAATAPVGINMRSGMQQQITPOPPNLAQMLAQORELYSQQRQRLIQOQRAMLRMQOS 765
QY 443 FANN-PPAS-----PAENGVCFFPQ-----BPPSGNHQQLIPVMSPSSEG 481
Db 766 FGNLPLPSSGLPVQTNPRLPQGAFOQFPYPNYGNTNPTGPAS-----TSPFSG 815
QY 482 LIYKPHPGMAH-----TGHYGGYGYHYMPTPMVMPQY-HPGMGFPFPPGNGYFPP-- 529
Db 816 LAANPEASLANRNSMVSRGTMGNIGQFGTGINPQMQQNVFOYFGAGVMPQGEANFAPSL 875
QY 530 ---YGMPTIMNPPYSSQQQQ-----QQQPMQNMQFQHGPNLQNTQQQQQORSNEPAP 580
```

[illegible]

581 QQ 582
—
933 AO 934

635-12
12, Application US/09125635
O. 6562589
INFORMATION:
NNT: THE UNITED STATES OF AMERICA represented by THE SE
F INVENTION: AIB1, A novel steriod receptor co-activator
REFERENCE: 43944
APPLICATION NUMBER: US/09/125,635
FILING DATE: 1998-08-21
APPLICATION NUMBER: 60/049,728
FILING DATE: 1997-06-17
OF SEQ ID NOS: 12
Patent In Ver. 2.0

; TYPE: PRI
 ; ORGANISM: Drosophila melanogaster
 ITS-09-506-066E-2

Query Match	4.1%;	Score 150.5;	DB 4;	Length 928;
Best Local Similarity	17.9%;	Pred. No. 0.00036;		

QY 109 RSSAQHDQRKMVREEEDFAVPVYINRRSQSHGRKSGIEKEK--HTPMV-APS----- 159

DB 307 KAKHAKRANALINDDLE
Qy 160 ---SH---SIRFEVNOTGSKO-----NYCLATCSKPEVRDQVKAN 195

```

QY
196 ARSGFVSLDVSVEEIDLEKSASHDRVDYNDYNASLRQE-----SRNR 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 240 LYRGG-KTRLKDTNGAES-----HLATENHSQEGHSGSPEDINDREYKSRACASLQQ 293

```
QY      294 INEEASDDV-----SDDSMDVDTSSSIDVSPDDVVGILGKRFWRAR-----KAIAA 339
```

```

340 QQRVFAV--QLFELHRLIKVQKLIASAFLDLELSIFGKVSANSIFVAVLQDGLZAK
      : : : : | | | | | | | | | | | | | | | | | | | | | | | |
QY

551 RNDLSIIRSEMEKNRLC-----FQLNGKPGQANVSPIQ-----584
      : : : : | | | | | | | | | | | | | | | | | | | | | | | |
Dh

```

[illegible]

Db 624 SPPTAEPPTS---IPAAPPAIEVNGQOHHPTHP--THPSHIN---HHENPQPHI-PIY 673

Db

674 HQQLAINPAVLAAQQTHTNTAHNKNLNGYDSFLHATICGGGAAAHSPPTATFSNVATVQPI 733

Db

674 HQQLAINPAVLAAQQTHTNTAHNKNLNGYDSFLHATICGGGAAAHSPPTATFSNVATVQPI 733

db 734 PKSQKNILQGVORLEQSQQQQQQRSSKDYKNYGNLIYAKLSEQLQKQKUREKQRKH- 750
-770 h n o o c c c c c t m y s v b b - - a p k s r o g s t g s p s g p o g i s g s k s f r p p f a a v d e d s i n n a p e o t 636

D6
793 --QQQHQLQDQFKDASKSEKPFISNSSSAGSALIGDAVTCIAKESBEEF--


```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: mouse
US-09-514-247A-8

Query Match      4.1%; Score 150.5; DB 4; Length 2441;
Best Local Similarity 24.3%; Pred. NO. 0.0014; Indels 95; Gaps 11;
Matches 75; Conservative 24; Mismatches 115;

QY 353 RLTKVKQKLTAAAPULLLDEISFLGKUSA--KSYVPVKLLPSEFLVKKPPLPHVVVVKORGDS 410
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2053 RMFVQPNRSISFSAQLDLRLTLKSPSPQQQQQVNLILKS---NPQLMAAFIKOR--- 2105

QY 411 EKTQHKMESSAENVVGRLSNQGHQOSNTMPANNPPASPAPNGYCFPPQPPSPSGNHQQ 470
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2106 -----TAKYV--ANQPMQPPQGLQSQPMQPPQGMHQQ 2137

QY 471 WLIPVMSPSGLIVKHPHGMNHTGHYGGYGHYMPMPVMVMPQYHPGMGFPPPCNGYFPPY 530
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2138 -----PSLQNLNAMQAG-----VPRGVFP-----PQPMGGILNPQ 2168

QY 531 GMMPTTINP-----YCSSQQQQQQQQQNEQMNQFHPGNLQ-N 566
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2169 GQALNIMNPCHNPNWNTNPNPYREVMVRRLQLHQQQQQQQQQQQQQQONSASLAGMGH 2228

QY 567 TQOQQRSDNEPAPQQQQQFTKSYPRARKRQGSTGSSPSGPQGISGSKSFRPFPAAVDE 626
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2229 SQQQPPQPGGYAPAMQQQRMQHLPQGSMSGMQ-AAPWGLQGMG---QFGLGADST 2283

QY 627 SNINNAPEQ 635
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2284 PNIQQALQQ 2292

RESULT 24
US-09-686-316-2
; Sequence 2, Application US/09686316
; Patent No. 6646115
; GENERAL INFORMATION:
; APPLICANT: Monemidy, Marc R.
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
; FILE REFERENCE: SALK1650-1
; CURRENT APPLICATION NUMBER: US/09/686,316
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US/08/961,739
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: US 194,468
; PRIOR FILING DATE: 1994-02-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: Mus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2441)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-686-316-2

```

	Query Match	4.1%;	Score 150.5;	DB 4;	Length 2441;
	Best Local Similarity	24.3%;	Pred. No. 0.0014;		
	Matches	75;	Conservative	24;	Mismatches 115; Indels 95; Gaps 11;
QY	353	RLIKVQKLIASPDLLIDETSFGLGKVA--KSPVVKLLPSEFLVPPLPHVVVKGQDS	410		
	:	:	:	:	:
	:	:	:	:	:
Db	2053	RMENVQNRISPSALQDLRLTKSPSSQQOQVLNILKS----NPQLMAAFIKOR---	2105		
	:	:	:	:	:
QY	411	EKTQDHQWESSAENVVGRLSNQGHHQSQNWYPFANNPPASPAPNGYCFPPQPDPSPGNHQQ	470		
Db	2106	-----TAKTV---ANQGMGPQPGLSQPGMQPQPGMHQQ	2137		

```

471 471 WLIVMSPSEGLIYKPHKMAHTGHYGYHYMPTMMPQYHPGCMGPPPGNGYPPY 530
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
2138 -----PSLQNLINAMQAG-----VPRPGVPP-----POPAMGGLNPQ 2168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
531 GMMPTIMNP-----YCSSQQQQQQQPNQMNQFHPGNLQ-N 566
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2169 GQALNIMNPHNPNMTNMMPQYEMVRRQLLHQQQQQQQQQQQQQQCNSSASLAGGMAGH 2228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
567 TOQQQRSDNEPAPOQQQKTSYPRARKSRQSGTSGSPGQIGSKGFRPPAAVDE 626
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2229 SQFQQQGGGYAPAMQQRMRQHLPTQGSSMGQM-AAPMQLGCMG---QPLGLGDST 2283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
627 SNINNAPBQ 635
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2284 PNIQALQQ 2292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 25
US-09-489-039A-12097
; Sequence 12097, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12097
; LENGTH: 1253
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-12097

```

Query Match	4.0%;	Score 148.5;	DB 4;	Length 1253;
Best Local Similarity	22.7%;	Pred. NO. 0.00082;	Indels 149;	Gaps 25;
Matches 102;	Conservative	37;	Mismatches 162;	
334	QY	334	KAIAANOQRVFA-----VQLFELHRLIKVKLI---AASPLLILDEISFLGKVS	379
Db		92	RLAARKEKLAEKFTNPMGRKTDAAALFSGRKMDGEEVQYSASGAPVAADDVLFSGASA	151
QY		380	AKSYPVVKLLPLPSBELVKPPLPHVVVQKRGDSEKTDQ-----HKM-----ESS	421
Db		152	ARPAEDDVLFSGASAVRP-----GDFDPYDLLNGHSGIAEPYSAAAAATAAPQAW	201
QY		422	AENVUGRLSNQGH-----QOSNMPFANNPPASPAPNGY	456
Db		202	ARSPV-----GHGGAAPAVQPEASYPPOQAYQPEAPFQAAAYQP---PAGOTAPQAY	251
QY		457	CEPPQPPSGNHQOOWLIPVMSPEG-----LIYKPHPCMAHTGHVGGYVGH-----YM--P	505
Db		252	--OPEPAP---YQQ-----PVYDPRVGOPAPQAYQPEPAPYQOPAYDPYAGQOPAPQAYQEP	303
QY		506	TPWVMPQYHPGMPFPPPGNGYFP---PYGMPTTIMNPFYCSQQQQQQOQPNQMNOFHGHP	562
Db		304	AFYQQPAYDPYAGQAP-PAYQPEPAPY-QOFT-YDPHAGQAPQAYQP-----EPA	352
QY		563	NLQNT-----QQQQQRSDNEPAPQQ-----QOFTKSYPRARKSGRSGTSSG	604
Db		353	PYQQPAPDPYAGQAPQATQYQPEPAPYQQPAYDPHAGQAPQAYQPEPAPYQPSQAFVPPP	412
QY		605	PSGPQGISGSKSRP---FAAVDDSDSNINN-----APEQTMTTTTTTTTTTTTQ	650
Db		413	EPPEVVQEEVKRPFLYYFEEVEEKARERELLASWYQPIPEPESPIATKPLTPPTTASK	472
QY		651	TTRDGGGVTRVIKVP---HNKLASENAA	677
Db		473	PPVE-----TTVISAVAAGVHQATASGGAA	498

Search completed: July 29, 2004, 08:30:42
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 08:23:21 ; Search time 50 seconds
(without alignments)

4360.193 Million cell updates/sec

Title: US-10-719-885-2

Perfect score: 3669

Sequence: 1 MKRGDEEKLEKMPFRLHV.....AARIFQIERKRYDSSKP 695

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3669	100.0	695	9	US-09-746-801A-2
2	3669	100.0	695	9	US-09-746-801A-35
3	2609.5	71.1	653	9	US-09-746-801A-13
4	904	24.6	324	9	US-09-746-801A-61
5	777	21.2	599	9	US-09-746-801A-59
6	773.5	21.1	760	9	US-09-746-801A-27
7	773.5	21.1	760	12	US-10-437-963-136891
8	725.5	19.8	534	16	US-10-424-599-230456
9	686.5	18.7	781	16	US-10-437-963-131862
10	621.5	16.9	389	9	US-09-746-801A-23
11	495.5	13.5	168	9	US-09-746-801A-63
12	459.5	12.5	317	9	US-09-746-801A-57
13	436.5	11.9	540	9	US-09-746-801A-33
14	397	10.8	248	9	US-09-746-801A-29
15	393.5	10.7	335	12	US-10-424-599-222485

16	377	10.3	334	12	US-10-425-114-42323	Sequence 42323, A
17	316	8.6	383	12	US-10-425-114-63567	Sequence 63567, A
18	306	8.3	213	12	US-10-424-599-253615	Sequence 253615, A
19	295	8.0	203	12	US-10-424-599-280208	Sequence 280208, A
20	281.5	7.7	179	9	US-09-746-801A-24	Sequence 24, Appl
21	256.5	7.0	110	12	US-10-424-599-285624	Sequence 285624, A
22	255.5	7.0	89	12	US-10-424-599-247729	Sequence 247729, A
23	231	6.3	49	9	US-09-746-801A-25	Sequence 25, Appl
24	212	5.8	208	12	US-10-425-114-43416	Sequence 43416, A
25	202.5	5.5	185	9	US-09-746-801A-19	Sequence 19, Appl
26	195	5.3	795	12	US-10-424-599-174901	Sequence 174901, A
27	193.5	5.3	802	15	US-10-369-493-22462	Sequence 22462, A
28	189.5	5.2	1130	14	US-10-032-585-7758	Sequence 7758, Ap
29	189	5.2	116	9	US-09-746-801A-16	Sequence 16, Appl
30	188	5.1	115	9	US-09-746-801A-47	Sequence 47, Appl
31	187	5.1	522	12	US-10-425-114-62797	Sequence 62797, A
32	186.5	5.1	1138	14	US-10-074-475-194	Sequence 194, App
33	181	4.9	631	12	US-10-424-599-266585	Sequence 266585, A
34	180	4.9	42	9	US-09-746-801A-37	Sequence 37, Appl
35	176.5	4.8	1132	12	US-10-425-114-62671	Sequence 62671, A
36	176.5	4.8	1291	12	US-10-312-352-32	Sequence 32, Appl
37	173	4.7	618	12	US-10-424-599-174900	Sequence 174900, A
38	172.5	4.7	2321	12	US-10-193-874-17	Sequence 17, Appl
39	172.5	4.7	2358	12	US-10-193-874-14	Sequence 14, Appl
40	172.5	4.7	2540	12	US-10-193-874-16	Sequence 16, Appl
41	172.5	4.7	2552	12	US-10-193-874-15	Sequence 15, Appl
42	170.5	4.6	189	9	US-09-746-801A-31	Sequence 31, Appl
43	167.5	4.6	1424	16	US-10-437-963-160683	Sequence 160683, A
44	166.5	4.5	1243	16	US-10-408-765A-1964	Sequence 1964, Ap
45	166.5	4.5	1441	15	US-10-355-218-2	Sequence 2, Appl
46	165	4.5	780	9	US-09-770-689A-5	Sequence 5, Appl
47	163	4.4	1965	15	US-10-369-493-3279	Sequence 3279, Ap
48	162	4.4	407	12	US-10-425-114-59366	Sequence 59366, A
49	162	4.4	407	16	US-10-437-963-147959	Sequence 147959, A
50	162	4.4	717	12	US-10-425-114-58581	Sequence 58581, A
51	162	4.4	816	14	US-10-207-706-3	Sequence 3, Appl
52	161	4.4	971	14	US-10-093-524-8	Sequence 8, Appl
53	161	4.4	2017	12	US-10-114-270-86	Sequence 86, Appl
54	160.5	4.4	1441	14	US-10-170-682-3	Sequence 3, Appl
55	160	4.4	699	16	US-10-437-963-192225	Sequence 192225, A
56	159	4.4	1272	16	US-10-437-963-191304	Sequence 191304, A
57	158.5	4.3	459	14	US-10-032-585-7777	Sequence 7777, Ap
58	158.5	4.3	1494	12	US-10-381-247B-2	Sequence 2, Appl
59	157.5	4.3	326	14	US-10-029-386-32987	Sequence 32987, A
60	157	4.3	1586	12	US-10-282-122A-49391	Sequence 49391, A
61	156	4.3	531	9	US-09-925-300-1444	Sequence 1444, Ap
62	156	4.3	1403	13	US-10-108-605-93	Sequence 93, Appl
63	156	4.3	1450	12	US-10-381-247B-17	Sequence 17, Appl
64	155.5	4.2	1036	9	US-09-842-256-3	Sequence 3, Appl
65	155.5	4.2	1367	15	US-10-320-797-3355	Sequence 3355, Ap
66	155	4.2	1644	16	US-10-408-765A-919	Sequence 919, App
67	154.5	4.2	1402	14	US-10-379-616-12	Sequence 12, Appl
68	153.5	4.2	1381	14	US-10-032-585-7784	Sequence 7784, Ap
69	153	4.2	274	9	US-09-850-887-4	Sequence 4, Appl
70	153	4.2	883	16	US-10-437-963-180574	Sequence 180574, A
71	152.5	4.2	1101	14	US-10-177-293-106	Sequence 106, App
72	151.5	4.1	756	15	US-10-104-047-2505	Sequence 2505, Ap
73	151.5	4.1	799	15	US-10-034-749-2342	Sequence 2342, Ap
74	151	4.1	626	12	US-10-425-114-60327	Sequence 60327, A
75	150.5	4.1	492	16	US-10-408-765A-1141	Sequence 1141, Ap
76	150.5	4.1	827	14	US-10-112-372-4	Sequence 4, Appl
77	150.5	4.1	1313	16	US-10-408-765A-2180	Sequence 2180, Ap
78	150.5	4.1	2441	13	US-10-109-886-8	Sequence 8, Appl
79	150	4.1	1572	15	US-10-116-275-179	Sequence 179, App
80	150	4.1	1791	16	US-10-389-566-2372	Sequence 2372, Ap
81	149.5	4.1	620	12	US-10-425-114-62454	Sequence 62454, A
82	149.5	4.1	643	14	US-10-192-985-1	Sequence 1, Appl
83	149	4.1	479	12	US-10-425-114-43565	Sequence 43565, A
84	148.5	4.0	813	14	US-10-112-372-6	Sequence 6, Appl
85	148.5	4.0	1566	16	US-10-437-963-106955	Sequence 106955, A
86	148	4.0	1001	15	US-10-262-445-66	Sequence 66, Appl
87	148	4.0	1005	13	US-10-114-091-2	Sequence 2, Appl
88	148	4.0	1005	15	US-10-262-445-62	Sequence 62, Appl

Thu Jul 29 08:52:24 2004

89 148 4.0 1068 15 US-10-262-445-64 Sequence 64, Appl
90 148 4.0 1446 12 US-10-343-649-7 Sequence 7, Appl
91 147.5 4.0 425 12 US-10-425-114-57692 Sequence 57692, A
92 147.5 4.0 427 12 US-10-425-114-45447 Sequence 45447, A
93 147.5 4.0 759 12 US-10-282-122A-71823 Sequence 71823, A
94 147.5 4.0 1360 16 US-10-473-574-12 Sequence 12, Appl
95 147 4.0 589 12 US-10-424-599-251769 Sequence 251769, A
96 147 4.0 976 15 US-10-144-194A-92 Sequence 92, Appl
97 146.5 4.0 638 14 US-10-038-010-4 Sequence 4, Appl
98 146.5 4.0 793 14 US-10-112-372-8 Sequence 8, Appl
99 146.5 4.0 1206 15 US-10-085-198-144 Sequence 144, App
100 146 4.0 1267 12 US-10-092-900A-128 Sequence 128, App

ALIGNMENTS

RESULT 1
US-09-746-801A-2
; Sequence 2, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-746-801A-2

Query Match 100.0%; Score 3669; DB 9; Length 695;
Best Local Similarity 100.0%; Pred. No. 1.7e-271;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRGKDEEKLPMFRLHVNDADKGGPRAPPNNKMALEYQLSIPSGRFGDHGTMSRSN 60
DB 1 MKRGKDEEKLPMFRLHVNDADKGGPRAPPNNKMALEYQLSIPSGRFGDHGTMSRSN 60

QY 61 NTSTLVHPGSSQPCGVERNLSVQHLDSAAQAOTEKFSVQMSFMENVRSSAQHDQKMY 120
DB 61 NTSTLVHPGSSQPCGVERNLSVQHLDSAAQAOTEKFSVQMSFMENVRSSAQHDQKMY 120

QY 121 REEDFAVPVYINRSRQSHGRKSGIEKEKHTPMVAPSSHHSIRFOEVNQTGSKQNVCL 180
DB 121 REEDFAVPVYINRSRQSHGRKSGIEKEKHTPMVAPSSHHSIRFOEVNQTGSKQNVCL 180

QY 181 ATCSKPEVRQVKANARSGGFVSLDVSVTEEDLEKSASHDRVNDYNASLRQESNRLL 240
DB 181 ATCSKPEVRQVKANARSGGFVSLDVSVTEEDLEKSASHDRVNDYNASLRQESNRLL 240

QY 241 YRDGKTRLKDTONGAESHLATENHSGHGSPEDINDREYSKSRACASLQQINEASD 300
DB 241 YRDGKTRLKDTONGAESHLATENHSGHGSPEDINDREYSKSRACASLQQINEASD 300

QY 301 DVSDSDMVDSISSIDVSPDDVVGILGQKRFWRARAKAIAQORFAVOLFEHLRLIKVQKL 360
DB 301 DVSDSDMVDSISSIDVSPDDVVGILGQKRFWRARAKAIAQORFAVOLFEHLRLIKVQKL 360

QY 361 IAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVKQGDSEKTDQHKMES 420
DB 361 IAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVKQGDSEKTDQHKMES 420

QY 421 SAENVVGRLSNQGHQHSNYMPFANNPPASAPNGYCFPPQPPSPGHNHQQWLIIPVMSPE 480
DB 421 SAENVVGRLSNQGHQHSNYMPFANNPPASAPNGYCFPPQPPSPGHNHQQWLIIPVMSPE 480

QY 481 GLIYKPHGMAHTGHYGGYGHYMPYTPMVMQYHPGMGFPFPPGNGYFPYGMPTIMNPPY 540
DB 481 GLIYKPHGMAHTGHYGGYGHYMPYTPMVMQYHPGMGFPFPPGNGYFPYGMPTIMNPPY 540

DB 481 GLIYKPHGMAHTGHYGGYGHYMPYTPMVMQYHPGMGFPFPPGNGYFPYGMPTIMNPPY 540
QY 541 CSSQQQQQQQNEQNNQFCHGNLQNTQQOQORSNDEPAPQOQQOQPTKSYPRARKSROGS 600
DB 541 CSSQQQQQQQNEQNNQFCHGNLQNTQQOQORSNDEPAPQOQQOQPTKSYPRARKSROGS 600
QY 601 TGSSPSGQGISGSGSRFPFAAVDEDSNINNAPEQTMTTTTTTTTQTTRDGGGVTR 660
DB 601 TGSSPSGQGISGSGSRFPFAAVDEDSNINNAPEQTMTTTTTTTTQTTRDGGGVTR 660
QY 661 VIKVYPHNAKLASENAARIFQSIQERKRYDSK 695
DB 661 VIKVYPHNAKLASENAARIFQSIQERKRYDSK 695

RESULT 2
US-09-746-801A-35
; Sequence 35, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 35
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-746-801A-35

Query Match 100.0%; Score 3669; DB 9; Length 695;
Best Local Similarity 100.0%; Pred. No. 1.7e-271;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRGKDEEKLPMFRLHVNDADKGGPRAPPNNKMALEYQLSIPSGRFGDHGTMSRSN 60
DB 1 MKRGKDEEKLPMFRLHVNDADKGGPRAPPNNKMALEYQLSIPSGRFGDHGTMSRSN 60

QY 61 NTSTLVHPGSSQPCGVERNLSVQHLDSAAQAOTEKFSVQMSFMENVRSSAQHDQKMY 120
DB 61 NTSTLVHPGSSQPCGVERNLSVQHLDSAAQAOTEKFSVQMSFMENVRSSAQHDQKMY 120

QY 121 REEDFAVPVYINRSRQSHGRKSGIEKEKHTPMVAPSSHHSIRFOEVNQTGSKQNVCL 180
DB 121 REEDFAVPVYINRSRQSHGRKSGIEKEKHTPMVAPSSHHSIRFOEVNQTGSKQNVCL 180

QY 181 ATCSKPEVRQVKANARSGGFVSLDVSVTEEDLEKSASHDRVNDYNASLRQESNRLL 240
DB 181 ATCSKPEVRQVKANARSGGFVSLDVSVTEEDLEKSASHDRVNDYNASLRQESNRLL 240

QY 241 YRDGKTRLKDTONGAESHLATENHSGHGSPEDINDREYSKSRACASLQQINEASD 300
DB 241 YRDGKTRLKDTONGAESHLATENHSGHGSPEDINDREYSKSRACASLQQINEASD 300

QY 301 DVSDSDMVDSISSIDVSPDDVVGILGQKRFWRARAKAIAQORFAVOLFEHLRLIKVQKL 360
DB 301 DVSDSDMVDSISSIDVSPDDVVGILGQKRFWRARAKAIAQORFAVOLFEHLRLIKVQKL 360

QY 361 IAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVKQGDSEKTDQHKMES 420
DB 361 IAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPPLPHVVVKQGDSEKTDQHKMES 420

QY 421 SAENVVGRLSNQGHQHSNYMPFANNPPASAPNGYCFPPQPPSPGHNHQQWLIIPVMSPE 480
DB 421 SAENVVGRLSNQGHQHSNYMPFANNPPASAPNGYCFPPQPPSPGHNHQQWLIIPVMSPE 480

QY 481 GLIYKPHGMAHTGHYGGYGHYMPYTPMVMQYHPGMGFPFPPGNGYFPYGMPTIMNPPY 540
DB 481 GLIYKPHGMAHTGHYGGYGHYMPYTPMVMQYHPGMGFPFPPGNGYFPYGMPTIMNPPY 540

Thu Jul 29 08:52:24 2004

Db 567 -----ARVIVKVPNHARSATESVARIFOSIQOER 595

RESULT 6

US-09-746-801A-27

; Sequence 27, Application US/09746801A

; Patent No. US20020083494A1

; GENERAL INFORMATION:

; APPLICANT: Wagner, et al.

; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM

; FILE REFERENCE: 1505-54357

; CURRENT APPLICATION NUMBER: US/09/746,801A

; CURRENT FILING DATE: 2000-12-20

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 27

; LENGTH: 760

; TYPE: PRT

; ORGANISM: Oryza sativa

US-09-746-801A-27

Query Match 21.1%; Score 773.5; DB 9; Length 760;

Best Local Similarity 32.1%; Pred. No. 3.9e-50;

Matches 261; Conservative 105; Mismatches 257; Indels 191; Gaps 36;

QY 4 GKDEE-KILEMPFRLHVNADK-GGPRAPPRKMALEYQLSIPSORF-----GDHGTMS 57

Db 12 GKEAKGVKMGPLFRLHVNDAAGGGPRAPPRKMALEYQFTVPSHRFSGGGGGVGG 71

QY 58 RSNTSTLVHPGSSQPCGVERNL-----SVQHLDSAAQAOTEKVFQSMF 104

Db 72 PAHSTSAASQSQSVYGRDSSLFQFNVPSNRPBGHSTKINSKINKISGRKELGM 131

QY 105 MEN-----VRSAQHDO-----KVVREEDFAPVYVINSRRSQSHGRTK 144

Db 132 LSSQTKGMDIYASGRSTAEAPQRAENTIKSSSKRLADDEFVWPSVFNRRFPQYSTQEN 191

QY 145 SGIEKEKHTPMWAPSSHHSIRFQVNVQTSQNV-CLATCSKPEVRDQV-KANARSGGV 202

Db 192 AGVQ-DQSTPLVANPHKS-----PSTVSKSTKCVNTVSKLERIHSVDSKSRTELKD 244

QY 203 ISLDVSVT-BEIDLEKASSHDRVNDYNASLRQESRN-RLYRDGKTR-LKTD-NGA 256

Db 245 KEMEAQTSKNVEVEKSSSFHASKDMF-----ESRHKVYPMKDKTGIINDSEPHGHN 298

QY 257 ESHLATENHS-----QEGHGSPEIDNDREY-----SKSRACASLOQIN 295

Db 299 SGHQATSRNGGSMKFPMPMRNEISSNSENDRHYNLPQGGIEETGKRLLEQHD 358

QY 296 EASDDVYS-----DDSMVDSISSIDVSPDDVVVILGQKRFWARKAIANQQ 341

Db 359 AEKSDVSRLLLEQHDENIDDDVSDSVRECITGWEISPKIVGAIGTKHFWKARRAIMNQ 418

QY 342 RVFAVQLFELHRLIKVOKLIAASPDLLDLDEISFLGKVSASYPVKLLPSEFLVKPLPH 401

Db 419 RVFAVQVFLHKLKVKQKLIASAPHLVIESDPCIGNLLGS-----KNKLVEENLKAQPLL 475

QY 402 VVVQKRGDSEKTDQHKWESSAENV-----VRLSNQGHQOOSNYMPAN---NPDA 449

Db 476 ATI-----DDVPSIQQPEVSKENTEDSPSPHDTGLSGQDQDAATNGSVKSNRRATPA 531

QY 450 SP-APNGCYCFPPQPPSGNHOQWLIPIVMSPEGLIYKPHG-----M 490

Db 532 SDNKNWGVQLQPP-----QNLVLPVMSPLGLVYKPYSGPCPAGSILAFYANCTPL 587

QY 491 AHTGHYGYHYHMTPTMWMQYHPGMPFP--PBGYFPPYGMPTIMNPYCSSQOQO 548

Db 588 SLPSTAGDFMNSAYGVPMHPQPHMGAPGPPSPMPN-YFPFESI--PVNPP-----635

QY 549 QCPNEQMOFGHGNLQ--NTQQOQSRSDNEPAPQOQOQPTKSYF---RARKSR-----597

Db 636 TAPAPVVSQGRHPSMPQYGNFEQOQSWISCN-----MSHPSGIWRPHASRDEA 684

QY 441 MPFANNPASPAPNGCYCFPPQ-PPPSGNHQ 469

Db 298 MPFASNPPAA---NGCYTTPPQHPPPSGGNQ 324

RESULT 5

US-09-746-801A-59

; Sequence 59, Application US/09746801A

; Patent No. US20020083494A1

; GENERAL INFORMATION:

; APPLICANT: Wagner, et al.

; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM

; FILE REFERENCE: 1505-54357

; CURRENT APPLICATION NUMBER: US/09/746,801A

; CURRENT FILING DATE: 2000-12-20

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 59

; LENGTH: 599

; TYPE: PRT

; ORGANISM: Lycopersicon esculentum

US-09-746-801A-59

Query Match 21.2%; Score 777; DB 9; Length 599;

Best Local Similarity 33.4%; Pred. No. 1.5e-50;

Matches 232; Conservative 95; Mismatches 216; Indels 152; Gaps 29;

QY 45 PSORFGDHGTMSRNNTSLVHPGSSQPCGVERNL-----LSVQHLDSAAQAOTEKVFQSM 102

Db 1 PSQRF-NSGVLPDLPNNTSKAPPS-SSQSGHDSGLPIQHPERRLADKPPGHSSDP 58

QY 103 SEMENVRSSAQHQKVMREEDFAPVYVINSRRSQSHGRTKSGIEKEKHTP---MVAPS 159

Db 59 STL-----LQVELKKRTEDEDTFPIFVNSKLQAHG--SHNVNMEKLSPSQQLPCPN 110

QY 160 -----SHHSIRFQVNVQTSQNVCLATCSKPEVRQVQVANARSGGVISLDV---SVT 210

Db 111 KELEGVTHLTRQQRNSQ--NKENL---KCTLAR-REKTTNSASKECRLDPQVGCSSIP 164

QY 211 BEIDLEKASSHDRVNDYNASLRQESRNRLYRDGKTRLDKTDNGAESHLATENHSQEGH 270

Db 165 EPVKGTGDSYPRKVFSEQL-----186

QY 271 GSPEDIDNDREYSKSRACASLIQINEASDDVSDSDSMVDSISSIDVSPDDVVVILGQKRF 330

Db 187 -TANDLVNDTESQEDRAHKSLOTGNLDRGDDLSETSRVESISGTDISPDIVIGILKRF 245

QY 331 WRARKAIANQORVFAVQLFELHRLIKVOKLIAASPDLLDLDEISFLGKVSASYPVKLLP 390

Db 246 WKARRAIVNQORVFAVQLFELHRLIKVORLIAGSPNSLESLEDPALYLGK-PLKSSSIKRLP 303

QY 391 SEFLVKPLPLVHVYVQKRGDSEKTDQHKWESSAENVVGRLS---NQGHQOOSNYMPFANNP 447

Db 304 LDCIVRE--SOSVLKRRHDSK-PHFRMEHTAESNVGKASLSTVQNGSQLSHKPFSTP 360

QY 448 PASPAPNG-----YCFPPQPPSGNHOQWLIPIVMSPEGLIYKPHGMAHTG-----494

Db 361 LPTPTVNDNAGPWCFF--QPPG---HQWLLIPVMSPEGLVYKPPGPGFTSPICGSGPP 415

QY 495 -----HYGYYGYHYHMTPTMWMQYHPGMPF---PPPGNGYFPPYGMPTIMNPYCSSQOQ 547

Db 416 GSSPTMGNFFAPTYGVPAFNPHYQ-GMGVPPFAPPTGHGYFROYGM--PAMNPPIS--468

QY 548 QOQPEQNOFGHGP-----NLQNTQQO---QQRSDNEPAPQOQOQPTKSYR 592

Db 469 --TASEENQYTMPLGLOHGFSGVDDVNIQHODSSNVLNKKNENVDVRYQSTKDNEVQ 526

QY 593 ARKSRQGSTGSSPSGPGISGSKSFPPFAAVDEDSNNINAPQMTTITTTITTTITTTT 652

Db 527 A-----SSASSPIETAGRNMLSLFPTSPVTNDRDGSQAQCPDNP-----566

QY 653 RDGGGVTRVIVKVPNHAKLASENAAIRFOSIQOER 687

```

QY 598 QGSGSPGSPGQIGSGSKFRPFAAEDSDSNINNAPEQMTTTTTTTTTTTTTRTTTQTTRDGGG 657
Db 685 QASASSPDRFCQSGS---GPVSAPFTVSAQNNQPOP-----SYSSRD--N 726
QY 658 VTRVIKVVPHNAKLASENAARIFQSIQERKRYD 691
Db 727 QTNVIKVVPHNSRTASESAARIFRSIQMERQRDD 760

RESULT 7
US-10-437-963-136891
; Sequence 136891, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 136891
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_38427C.1.pap
US-10-437-963-136891

Query Match 21.1%; Score 773.5; DB 16; Length 760;
Best Local Similarity 32.1%; Pred. No. 3.9e-50;
Matches 261; Conservative 105; Mismatches 257; Indels 191; Gaps 36;

QY 4 GKDEE-KILEMFFRLHVNDADK-GGPRAPPKMKALYQLSIPSORF-----GDHGTMS 57
Db 12 GHEAGKVMGPFPRLVHNDAAKGGPRAPPKMKALYEQFTVPSHRFSGGGGGVGS 71
QY 58 RSNNTSTLVHPGSPGCGVERNL-----SVQHLDSAAQAQATEKFVYSQMSF 104
Db 72 PAHSTSAASQSQSQVYGRDSSLFPNVPSNRPCHSTKINSKINKKISGRKEJGM 131
QY 105 MEN-----VRSAQHDQR-----KMVREEDFAPVYVYINRSRSQSHGRITK 144
Db 132 LSSQTKGMDIYASRTAEAPQRAENTIKSSSGKRLADDDEFPVPSVFNRSRFPQYSTQEN 191
QY 145 SGIEKEKHTPMVAPSSHSIRFQEVNQTGSKQNV-CLATCSKPEVRDQVK-ANARSGGFV 202
Db 192 AGVQ-DQSTPLVANPHKS-----PSTVSKSGSTKCYNTVSKKLEIRHVSDVSKSRPLKD 244
QY 203 ISLDVSVT-BEIDLKFSASSHRVDYNASLRQESRN-RLYRDGKTR-LKDTD---NGA 256
Db 245 KEMEAQTSKXNVEKSSSFHASKDMF-----ESRHAKVYPMKDTGTINDSDPFGHN 298
QY 257 ESHLATENHS-----QHGSGPEDINDREY-----SKSRACASLOQIN 295
Db 299 SGHQATSRNGSGMKFQNPFRNRNEISSNPSSSENTDRHYNLPQGGTEETGTRKRLLLEQHD 358
QY 296 EASDDVS-----DQSMVDSISIDVSPDDVGVILGQKFRWARKAIAANQ 341
Db 359 AEKSDVDRLELLEQHDAAENIDVDSVSEICITGWEISPKIIVGAIGTKHFWRKARRAIMNQ 418
QY 342 RVFAVQLFELHRLIKVQKLIASPDLLLEISFLGKVSAKSVYVKKLLPSEFLVKPPLPH 401
Db 419 RVFAVQVVELHKLKVKQKLIASPHVLIESDPLGNALLGS---KNKLUVEENLKAQPLLV 475

```

```

QY 402 VVKQRGDSEKTDQHKMESSAENV-----VGRLSNQGHQOOSNMPFAN---NPPA 449
Db 476 ATI-----DDVEPSLQQPEVSKENTEDSPPSPHDTGLSGQORDQAATNGSVKSNRRATPVA 531
QY 450 SP-APNGYCFPPPPSPSGNHQWLIPVMSPSSEGLIYKPHG-----M 490
Db 532 SDNKQNNWGVQLQPP-----QNQWLVPVMSPLEGLIYKPYSGPCPPAGSILAPFYANCTPL 587
QY 491 AHTGHYGGYGYHYMPTPMVMPQYHPGGMFPP--PGNGYFPPYGMMPMTIMNYPYSSQOQQQ 548
Db 588 SLPSTAGDFMNSAYGVPMHPQPCQHMGAAPGPPSPMPN-YFPFESI--PVMNP-----635
QY 549 QQPNQMNQFQGNLQ---NTQQOQORSNEAPQOQQOQPYKSY-----RARKSR-----597
Db 636 TAPAPVVEQGRHSPMPQPYGNFEQQSWISCN-----MSHPSGIWRFHARSDEA 684
QY 598 QGSGSPGSPGQIGSGSKFRPFAAEDSDSNINNAPEQMTTTTTTTTTTTTTRTTTQTTRDGGG 657
Db 685 QASASSPDRFCQSGS---GPVSAPFTVSAQNNQPOP-----SYSSRD--N 726
QY 658 VTRVIKVVPHNAKLASENAARIFQSIQERKRYD 691
Db 727 QTNVIKVVPHNSRTASESAARIFRSIQMERQRDD 760

RESULT 8
US-10-424-599-230456
; Sequence 230456, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 230456
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_50123C.1.pap
US-10-424-599-230456

Query Match 19.8%; Score 725.5; DB 12; Length 534;
Best Local Similarity 40.0%; Pred. No. 1.1e-46;
Matches 199; Conservative 64; Mismatches 157; Indels 77; Gaps 21;

QY 235 ESRNLXRDGG--KTRLKDTDNGAESHLATENHSQEHGSPEDINDREYSKSRACASLQ 292
Db 75 ESRSDIDGNGCLVESARETDKG---NAPTANQT-----SPADAINDTEHHDTRMGSPIQ 126
QY 293 QINEEASDDVSDSDVDSISIDVSPDDVGVILGQKFRWARKAIAANQRFVAVOLFELH 352
Db 127 RGNLNESSNASKISMVENLSTVRISPDVVGIIQKHEFWKARRAIAANQRFVAVOLFELH 186
QY 353 RLKVKQLIAASPDLLLEISFLGKVSAKSVYVKKLLPSEFLVKPPLPHVVKQGDSEK 412
Db 187 RLKVKQLIAASPDLLLEISFLGKVSPPKGTTPKK-LALEYVVKPRQON---LKRKDDSEK 243
QY 413 TDQHKMESSAENVVGRLS-----NQGHQOOSNMPFANP-----PASAPNGYCFPPPOP 463
Db 244 LN-HKMECSAENAVGKTSLSVKDGSH--LSKCTFPFGNQHTNVAAADSGMGFWCFNQSP 301
QY 464 PSQHQQWLIPVMSPSSEGLIYKPHGMAHTG--HYGG-----YYGHYMTPTMVMPOY 513
Db 302 ---GHFWLIPVMTFSEGLVYKPYGPGFTGTGCGGGGTGLREDDRRGGSFNNPGVIGPTS 357
QY 514 HPGMGFPP---PG-NGYFPPYGMMPMTIMNYPYSSQOQQOQQOQOQFGHGNLQNTQQ 569

```


Db 55 NISDSLSTFSLSLPPPPNNAKLLDGPENKQSPPIYNTKFKLKKNGKNGINYSKPGSSVTN 114
Qy 108 VRSSA--QHDQKQVREDEFAVPVINSRRSQSHGRTKSGIEKKEKHTPMVAPSSHHR 165
Db 115 TYPSSIKQNEVYKNTSLDSIKSPIVHSE-----IDPQANTDL-----SLQ 156
Qy 166 PQEVNQTGSKQNVCIATCKPVRQOVKANAKSGGFVSLDVSVTTEEIDLEKSASHDRV 225
Db 157 F-----CtSGSKP-----GGEAVVGSKILLSE-----RL 181
Qy 226 NDYNASLRQESNRRLYRGGKTLKDTDNGAESHLATENHSGEHCSPEDIDNDREYSKS 285
Db 182 EDEN-----QNGSFVWKTQSYRR-----NFAEFNETQ-KKP 213
Qy 286 RACASIQQINEBASDDVSDSDSVSSIDVSPDDVVGILGQKRFWRARAKAJANQORVFA 345
Db 214 KTLPREQV-----ASNCASATESLSGSIASSYDIARVIGREKFWKMTYMINQOKIPA 266
Qy 346 VQLFELHRLIKVKQLIAASPDLLLDLDEISFLGKVSAXSPVKLLPSEFLVKPPLPHVVVK 405
Db 267 GQVFELHRLIMVQKMAKSNLFLES-----K 293
Qy 406 QRGDSEKT--DQHKMESSAENVVGRLSNOGHQOQSNMPPFANNPPASAPNGY----- 456
Db 294 LNVGKHGTRSRSSHOLAASKV--RKPNTENHK-----PVPEEYPEHMKPK 337
Qy 457 -----CFPPQ-PPPSGNHQWLIIPVMSPSGLIYKPHGPMAGHTGYGYGH 502
Db 338 LPIPSISKELVTIWMQQLPPGN--QWLVPIVITDGLVYKPPG----- 382
Qy 503 YMTFP-----MVMQY-----HFGMGFP-----PPGNGYFPPYGMPTIMNPPYCSSQQQQQ 549
Db 383 --PCPSSSAFVFPVYQDSLETFFRPFPVSSPFSYFPPPNARTV----- 427
Qy 550 QPNEQMNQFHPGNLQNTQQOQORSNDNEPAPQOQOOP---TKSYPRARKSRQGSTGSSPS 606
Db 428 ---DQTNPPG-----QFQWSNTSSHMTCAIPFSLKKSQESNDSDIHGSTASSP- 473
Qy 607 GPOGISGSKFRPFAAVIDESNINNAPEQTMTTTRTTRTQT---RDGGGVTRVI 662
Db 474 -----PEKHLEVLPLFPTEPTHTQDYEKQKQOQMLRAI 507
Qy 663 KVPHPNKLASENNAARIFQIOERKRYD 691
Db 508 KAVPHNSTSASESAAIRIFQIOERDRSD 536

RESULT 14
US-09-746-801A-29
; Sequence 29, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746.801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (88)..
; OTHER INFORMATION:
; NAME/KEY: unsure
; LOCATION: (98)..
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
; NAME/KEY: unsure
; LOCATION: (113)..
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence

; NAME/KEY: unsure
; LOCATION: (114)..
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
US-09-746-801A-29
Query Match 10.8%; Score 397; DB 9; Length 248;
Best Local Similarity 29.5%; Pred. No. 4.8e-22;
Matches 121; Conservative 34; Mismatches 180; Gaps 13;
Qy 237 EASDDVSDSDSVSSIDVSPDDVVGILGQKRFWRARAKAJANQORVFAVQLFELHRLIK 356
Db 3 EQNDLSDSS--VESLPGMEISPDVVSAGIKPKHFKARRAIVNQORVFAVQVFLHRLIK 61
Qy 357 VQKLIASPDLLLDLDEISFLGKVSAXSPVKLLPSEFLVKPPLPHVVVKQRGDSEKTDH 416
Db 62 VQKLIASPHVLIEGDCGLGSLAVSXK----- 89
Qy 417 KMESAEENVVGRLSNOGHQOQSNMPPFANNPPASAPNGYCFPPQPPPSGNHQWLIIPV 476
Db 90 -----RLS-----QWLIIPV 99
Qy 477 SPSEGLIYKPHGPMAGHTGYGYGHYMPTEPMVQYHP-----GMGFPFPPGNG----- 525
Db 100 SPFEGILVYKPYKXPSG-----GSLAPPFFAS--YFTSSSTAGDGFMSACGARLM 151
Qy 526 ---YFPPYGMPTIMNPPYCSSQQQQOQPNBOMQFHPGNLQNTQQOQORSNDNEPAPQ 581
Db 152 SAPVTFPSFS-NEAV-----SGSAVEQVSHVA---ASQHKRNSCSEAVLA 192
Qy 582 QQQQPKSYPRARKSRQGSTGSPSPQGISGSKSPFAAVIDESNINNAPEQTMTT 641
Db 193 SRDSEV-----QGSASPSASSE----- 210
Qy 642 TITRTTITVOTTRDGGVTRVIVKVPHPNKLASENNAARIFQIOERKRYD 691
Db 211 ---TAAQ-----PVIRVVPHTARTASESAAIRIFPSIQWERKQND 247
RESULT 15
US-10-424-599-222485
; Sequence 222485, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 222485
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_42933C.1.pep
US-10-424-599-222485
Query Match 10.7%; Score 393.5; DB 12; Length 335;
Best Local Similarity 34.0%; Pred. No. 1.4e-21;
Matches 127; Conservative 49; Mismatches 116; Indels 81; Gaps 19;
Qy 358 QKLIASPDLLLDLDEISFLGKVSAXSPVKLLPSEFLVKPPLPHVVVKQRGDSEKTDH 417
Db 3 RKLIASPDVLEDDVAFMGKFPLETSEPKN--ISLEVVVEPQQQN--PKRKNDSKLN-HK 58
Qy 418 MESSAENVVGRLSN-----QGHQOQSNMPPFANNPP--PASAPNG-----YCFPPQPPPSGN 467
Db 59 TECSAENAVAKRTSPSPKNGSH-LANHTPFGTHQANVASDNKTSWCFNQTP----- 112

```
QY 468 HQWLIPVMSPEGLIYKPH--PGMAHTGHVGGYGHVMPPTMVMQVHPGMPFP-- 522
Db 113 GHOWLIPVMSPEGLIYKPH--PGMAHTGHVGGYGHVMPPTMVMQVHPGMPFP-- 171
QY 523 -----GNGYFPYPMPTTMMNFCSSQQQQQPNQMNQFHPGNL-QN----- 566
Db 172 VRVSPFVPELSHYFALGFM--RVVN-----QATSGSAVEQVNOFAAQSGHONGHSSV 223
QY 567 -----TQOQOORSDNEPAQOQOQPTKSYPRAKSR--QGSTGSSPSG-PQGISGSKSFR 618
Db 224 EGADFNTHNQSSNLVPQKNGARLHVKKSQALKERGLOQSTRSSPSEMAQGITRAGKI-- 281
QY 619 PFAAVDESDINNAPEQMTTTTTTITVITQITRDGGVTRVTKVPHNAKLAASENAAR 678
Db 282 -----ADGSDAQSLSHAVETROQ-----TOALKVVPHNKRSATESAAR 320
QY 679 IFQSIQERKRYD 691
Db 321 IFQSIQERKQHD 333

RESULT 16
US-10-425-114-42323
; Sequence 42323, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 42323
; SEQ ID NO 42323
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3061-038-G8_FLI.pep
US-10-425-114-42323

Query Match 10.3%; Score 377; DB 12; Length 334;
Best Local Similarity 33.4%; Pred. No. 2.5e-20;
Matches 124; Conservative 45; Mismatches 108; Indels 94; Gaps 16;

QY 192 VKANARSGGFVSLDVSVTEIDLE--KSASHDRVNDYNASLRQESRNLRYDGGKTRL 249
Db 25 VKSKGPSPG---IKEKEPVQVRIDLEDKETTPSFQVINDKTWSPDPKLSHSM-----DEL 75
QY 250 KPTDNGAESH-LATENHSGEGHSP-----EDIDND-----REYSKS 285
Db 76 KQQAHAESYQIRTNENAVETQSPKNGVSLLSKPYVDRREQNGSDLLHGLRETVK 135
QY 286 RACASLQOINEASDDVSDSDSISSIDVSPDDVVGILGQKRWARKALANQORVFA 345
Db 136 R---KRSHHDVEQNDLSDSS--VESLPGMEISPDVVSAIGPKHFWKARRALVNOQVFA 191
QY 346 VOLFEHLRLIKVQLIAASPDLLDEISFLGKVSAGKSPVKKLSEFLVKPPLPHVVVK 405
Db 192 VOVFEHLRLIKVQLIAASPHVLIBGDCPLGKSLAVS---KKRLAGD-----VE 237
QY 406 ORGDESEKTDQ-----HKMESSAENVVGRLSNQGHQOQ---SNYMPANPPASPAPNGYCF 458
Db 238 TQLESAKNDGVRPTQLESHSEKTEANQSPSPSQDEQAATNGDVAASMETPSDNKQKSWCI 297
QY 459 PPQPPSGNHQOWLIPVMSPEGLIYKPHGPGMAHTGHVGGYGHVMPPTMVMQVHPGPGM 518
Db 298 P--APPS-----QWLIPVMSPEGLVYKP-----YIGH----- 323
```

```
QY 519 FRPPGNGYFPP 529
Db 324 CPPVGSLLAPP 334
```

RESULT 17

```
US-10-425-114-63567
; Sequence 63567, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63567
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLTEOSINTE47H05_FLI.pep
US-10-425-114-63567

Query Match 8.6%; Score 316; DB 12; Length 383;
Best Local Similarity 28.4%; Pred. No. 1.4e-15;
Matches 118; Conservative 51; Mismatches 124; Indels 122; Gaps 18;

QY 1 MKRG--KDE---EKILEPMFPLHLVNDADKGGPRAPPRNKMAIYEQLSIPSORFGDHGTM 55
Db 11 MRGATKDDAAPDKVMGPLFPLHLVNDTLKGGPRAPPRNKMAIYEQFSVPESHRY----- 64
QY 56 NSRSNTSTLVHPGS-SOPCGVERNLSVQHLDLSAANO-----ATEKPV 99
Db 65 -----SAAVPPAPSPAPPWGAQRPASA--VPSTASQVGGGDRPIFPLFRVPSTEPVR 115
QY 100 SOMSFENVRSSAQ-----HDKRKWREEDF 126
Db 116 S--SDQTANSGQAANGTIAESWQROSTHLKSDTNAAGPPAGNNSVGKKLANDDDF 173
QY 127 AVPVYINRRSOSHGRKSGIEKHEKHTPMVAPSSHHSIRFQVNTGTGSKQNVCLATCSKP 186
Db 174 TVPSVLYSGMP-----PHSSQEKLTLPFTTSPCKSPAKYSSTDK-----RRL 216
QY 187 EVRD--QVKANARSGGFVSLDVSVTEIDLE--KSASHDRVNDYNASLRQESRNLRYR 242
Db 217 EGMDSADVSKSGPSG---IKEKEPVQVRIDLEDKETTPSFQVINDKTWSPDPKLSHSM-- 271
QY 243 DGGKTRLKTDTNGAESH-----ATENHSGEGH-----SPEDIDNDRYKSKRACASL 291
Db 272 ----DLKQQAHAESYQIRTNENAVETQSPKNGVSLLSKPYVDRREQNGSDLLHGL 327
QY 292 QQINE-----FASDDVSDSDSISSIDVSPDDVVGILGQKRWARKAI 337
Db 328 RETGEKRRSHHDVEQNDLSDSS--VESLPGMEISPDVVSAIGPKHFWKARRAI 381
```

RESULT 18

```
US-10-424-599-253615
; Sequence 253615, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
```

Thu Jul 29 08:52:24 2004

;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

;; FILE REFERENCE: 38-21(53223)B
;; CURRENT APPLICATION NUMBER: US/10/424,599
;; NUMBER OF SEQ ID NOS: 285684
;; SEQ ID NO 253615
;; LENGTH: 213
;; TYPE: PRT
;; ORGANISM: Glycine max
;; FEATURE:
;; OTHER INFORMATION: Clone ID: PAT_MRT3847_71038C.1.pep
US-10-424-599-253615

Query Match 8.3%; Score 306; DB 12; Length 213;
Best Local Similarity 39.7%; Pred. No. 3.5e-15;
Matches 81; Conservative 34; Mismatches 69; Indels 20; Gaps 8;
QY 1 MKRGDEEKILPEMPFRLHVNDAKGGPRAPPRNKMALYEQLSIPQSGDGHGTMSRSN 60
DB 1 MKRGDEEKILPEMPFRLHVNDAKGGPRAPPRNKMALYEQLSIPQSGDGHGTMSRSN 58
QY 61 NTSTLVHPGPSSQPCGVERN-LSVQHLDSAAQAQTEKFVSQMSFMENVRS--AQHDQ 116
DB 59 NSSNIVP-PPTQGNHRSYVYVRFSSOTPHRAESYISROS-DDGSRSTSLVQLER 116
QY 117 RKMVREEDFAPVYINRSRQSHGRKTSKIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQ 176
DB 117 RKVDDDD----IHVYTCSGIDQNDKMTESVDGKKTFPGARNFCYSV---AVQNDGDKD 169
QY 177 NVCLATCSKP-----EVRDQVKAN 195
DB 170 PTFSSCSLSPVLDLRKDVNGNEAN 193

RESULT 19
US-10-424-599-280208
;; Sequence 280208, Application US/10424599
;; Publication No. US20040031072A1
;; GENERAL INFORMATION:
;; APPLICANT: La Rosa Thomas J
;; APPLICANT: Kovalic David K
;; APPLICANT: Zhou Yihua
;; APPLICANT: Cao Yongwei
;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;; FILE REFERENCE: 38-21(53223)B
;; CURRENT APPLICATION NUMBER: US/10/424,599
;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 285684
;; SEQ ID NO 280208
;; LENGTH: 203
;; TYPE: PRT
;; ORGANISM: Glycine max
;; FEATURE:
;; OTHER INFORMATION: Clone ID: PAT_MRT3847_95050C.1.pep
US-10-424-599-280208

Query Match 8.0%; Score 295; DB 12; Length 203;
Best Local Similarity 38.9%; Pred. No. 2.3e-14;
Matches 82; Conservative 26; Mismatches 79; Indels 24; Gaps 5;
QY 1 MKRGDEEKILPEMPFRLHVNDAKGGPRAPPRNKMALYEQLSIPQSGDGHGTMSRSN 60
DB 1 MKRGDEEKILPEMPFRLHVNDAKGGPRAPPRNKMALYEQLSIPQSGDGHGTMSRSN 58
QY 61 NTSTLVHPGPSS-QPCGVERN-LSVQHLDSAAQAQTEKFVSQMSFMENVRS--AQHDQ 118
DB 59 NSSNIVPPTSTEGNRSYISROSDGARN-----TSVLQLERRK 102
QY 119 MYREEDFAPVYINRSRQSHGRKTSKIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV 178
DB 103 KV-----DDGIHVYTCRIDQNDKMLSPNGKLTFFPGARNFCYSVAVQNGGDKDTQFG 158

QY 179 CLATCSKPEVRDQVKANARSGGFVISLDSV 209
DB 159 FLPMRDXDARKNEANTHVSRSRQKLKLSV 189
RESULT 20
US-09-746-801A-24
;; Sequence 24, Application US/09746801A
;; Patent No. US20020083494A1
;; GENERAL INFORMATION:
;; APPLICANT: Wagner, et al.
;; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;; FILE REFERENCE: 1505-54357
;; CURRENT APPLICATION NUMBER: US/09/746,801A
;; CURRENT FILING DATE: 2000-12-20
;; NUMBER OF SEQ ID NOS: 68
;; SOFTWARE: Patent in version 3.1
;; SEQ ID NO 24
;; LENGTH: 179
;; TYPE: PRT
;; ORGANISM: Lycopersicon esculentum
US-09-746-801A-24

Query Match 7.7%; Score 281.5; DB 9; Length 179;
Best Local Similarity 41.1%; Pred. No. 2.1e-13;
Matches 79; Conservative 25; Mismatches 61; Indels 27; Gaps 9;
QY 1 MKRGDEEKILPEMPFRLHVNDAKGGPRAPPRNKMALYEQLSIPQSGDGHGTMSRSN 60
DB 1 MKRGDEEKILPEMPFRLHVNDAKGGPRAPPRNKMALYEQLSIPQSGDGHGTMSRSN 59
QY 61 NTSTLVHPGPSSQPCGVERNLSVQHLDSAAQAQTEKFVSQMSF-----MENVRSSAQHDQ 116
DB 60 NSANLVLPHP-SQENEHERGVLFSSR-QLPALRHPVEKPYGRSSGSNTPLREVKSKRQ--- 114
QY 117 RKMVREEDFAPVYINRSR-----RSQSHGRKTSKIEKE-----KHTPMVAPSSHHSIRFQ 168
DB 115 ----TEKEDFRVPTFDNSKERAVENTEDYSGTSDIDRDSSTLKRDTQOL----SHVTPREN 167
QY 169 VNQTGS--KQNV 178
DB 168 VNTFGSHKTN 179

RESULT 21
US-10-424-599-285624
;; Sequence 285624, Application US/10424599
;; Publication No. US20040031072A1
;; GENERAL INFORMATION:
;; APPLICANT: La Rosa Thomas J
;; APPLICANT: Kovalic David K
;; APPLICANT: Zhou Yihua
;; APPLICANT: Cao Yongwei
;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;; FILE REFERENCE: 38-21(53223)B
;; CURRENT APPLICATION NUMBER: US/10/424,599
;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 285684
;; SEQ ID NO 285624
;; LENGTH: 110
;; TYPE: PRT
;; ORGANISM: Glycine max
;; FEATURE:
;; OTHER INFORMATION: Clone ID: PAT_MRT3847_99948C.1.pep
US-10-424-599-285624

Query Match 7.0%; Score 256.5; DB 12; Length 110;
Best Local Similarity 65.0%; Pred. No. 8.4e-12;
Matches 52; Conservative 7; Mismatches 20; Indels 1; Gaps 1;
QY 1 MKRGDEEKILPEMPFRLHVNDAKGGPRAPPRNKMALYEQLSIPQSGDGHGTMSRSN 60

Db 1 MKRGKDEKVGMPFRLHVNDDTKGGPRAPPRNKMALYQFSIPSORF-NSGVLPLNPN 59

Qy 61 NTSTLVHPGPSSQCGVERN 80
Db 60 ISSNTVPPASSSLRTVPERN 79

RESULT 22

US-10-424-599-247729
; Sequence 247729, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 247729
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(89)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_65729C.1.pep
US-10-424-599-247729

Query Match 7.0%; Score 255.5; DB 12; Length 89;

Best Local Similarity 66.2%; Pred. No. 7.5e-12; Length 89;
Matches 53; Conservative 12; Mismatches 14; Indels 1; Gaps 1;

Qy 308 VDSISIDVSPDV-VGILGQKRFKARKAIANQQRVAVQVLFHLRLIKVKQLIAASPD 366
Db 1 VTNLSLLVSPDVVXGILGQKHLWKARKIANQQSFAVQVLFHLRLIKVQQLIAASPD 60

Qy 367 LLLDEISFLGKVSASKSPVK 386
Db 61 VLFEDGAFMGKYSLMGSPPK 80

RESULT 23

US-09-746-801A-25
; Sequence 25, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-746-801A-25

Query Match 6.3%; Score 231; DB 9; Length 49;

Best Local Similarity 85.7%; Pred. No. 2.4e-10; Length 49;
Matches 42; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MKRGKDEKILPEMPFRLHVNDDTKGGPRAPPRNKMALYEQLSIPSORF 49
Db 1 MKRGKGEKLMGMPFRLHVNDDTKGGPRAPPRNKMALYEQLSIPSORF 49

RESULT 24

US-10-425-114-43416
; Sequence 43416, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43416
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700237614_FLI.pep
US-10-425-114-43416

Query Match 5.8%; Score 212; DB 12; Length 208;

Best Local Similarity 25.5%; Pred. No. 5.2e-08; Length 208;
Matches 76; Conservative 34; Mismatches 72; Indels 116; Gaps 11;

Qy 411 EKTDOHKMESSAENVVGRLSNQGHCQSNMYMFPANNPPASPAPNGYCFPPQPPPSGNHOQ 470
Db 9 EXTENQPSQDE---QAATNGDVAAASHMTPSDNK-----QKSWCIP--APPS---Q 53

Qy 471 WLIPVMSPEGLIYKPHPGMAHTGHYGYGHMPTMVMPOYHPCMGFPFPNGYFPPY 530
Db 54 WLIPVMSPEGLIYKPHPGMAHTGHYGYGHMPTMVMPOYHPCMGFPFPNGYFPPY 530

Qy 531 GMPITMMPYCSSQQQQQQQQPNEQMGHPGNLQNTQQQQORDNEPAPQQQQPTKSY 590
Db 86 -----FASY 89

Qy 591 PEARKSROGST-GSSPSGPGTSGSKSRFPF-----AAVDEDSNI-----NNAPQ 635
Db 90 PFSSTTAGGDFMSSACGARLMSAPVYFSPSMPAVSGSAVEQVSHVAASQHKRNSCSEA 149

Qy 636 TMTTTTTTTR--TTVTQTRDGGVTRVKKVPHNAKLASENAARIFOSIOERKRYD 691
Db 150 VLASRDSEVQSSASSPASSSETAAQPRVIRVVPHTARTASEAARIFRSIQWERKQND 207

RESULT 25

US-09-746-801A-19
; Sequence 19, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Glycine max
US-09-746-801A-19

Query Match 5.5%; Score 202.5; DB 9; Length 185;

Best Local Similarity 33.3%; Pred. No. 2.4e-07; Length 185;
Matches 69; Conservative 20; Mismatches 69; Indels 49; Gaps 9;

us-10-719-885-2.rapb

Thu Jul 29 08:52:24 2004

QY 503 YMPTEWVMPQYHPGNG-----PPPPGNGYFPYGMPTIMNPHYCSSQQQQQQQPPNEQMNQF 558
Db 10 FWNPAYQFPASHPVVGVSPFVPFASHTYFAFGM--PVMN-----QATSGSAVEQVNOF 61
QY 559 GHPCNL-QN-----TQQQQORSNNEPAPQQQQQPTKSYPEARKSR--QGSGSSP 605
Db 62 AAQSGHGQNGHSSVEGADFNTHNQSSNLPVQKNGARLHVKKSOALKERGLQGSTRSSP 121
QY 606 SG-POGISGSKSFREFAAVDEDSNINNAPEQTMTTTTTTTTRTVTQTTRDGGGVTRVIKV 664
Db 122 SEMAQGIRARKI-----ADGSDAQSLSLHADETRQQ-----TQAIKV 158
QY 665 VPHNAKLASENAARIFQSIQERKRYD 691
Db 159 VPHNRKSATESAARIVQSIQERKQHD 185

Search completed: July 29, 2004, 08:32:20
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 08:20:25 ; Search time 14 seconds
(without alignments)
2584.911 Million cell updates/sec

Title: US-10-719-885-2

Perfect score: 3669

Sequence: 1 MKRGKDEKILEMPFRLHV.....AARFQSIQERKRYDSSKP 695

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141601 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141601

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3669	100.0	695	1	RLF3 ARATH
2	197	5.4	971	1	R3HD HUMAN
3	193.5	5.3	802	1	NAB3 YEAST
4	189.5	5.2	738	1	YKF4 YEAST
5	184	5.0	1048	1	SRA4 RAT
6	179	4.9	971	1	CLA4 CANAL
7	171.5	4.7	1596	1	MAM DROME
8	171	4.7	1157	1	SRA4 HUMAN
9	167	4.6	943	1	YMI1 YEAST
10	166.5	4.5	1429	1	EXPA DROME
11	165.5	4.5	1983	1	TF20 MOUSE
12	165	4.5	612	1	ADFI CANAL
13	164.5	4.5	629	1	PAB2 ARATH
14	164.5	4.5	2716	1	OSA DROME
15	163.5	4.5	1516	1	NCO2 XENLA
16	162.5	4.4	656	1	CPH1 CANAL
17	162	4.4	816	1	ATX1 HUMAN
18	162	4.4	3726	1	ABF1 MOUSE
19	161	4.4	1960	1	TF20 HUMAN
20	158	4.3	1089	1	Y553 HUMAN
21	158	4.3	2056	1	CBP1 CAEEL
22	156	4.3	1403	1	PROS DROME
23	155.5	4.2	964	1	YQKA SCHPO
24	154.5	4.2	1902	1	SMF1 HUMAN
25	153	4.2	1693	1	SAS DROME
26	152.5	4.2	734	1	YHU5 YEAST
27	152	4.1	418	1	SSXT MOUSE
28	152	4.1	668	1	PAB5 ARATH
29	152	4.1	1586	1	SN22 HUMAN
30	151.5	4.1	2195	1	SC16 YEAST
31	151	4.1	1082	1	NCO3 RAT
32	150.5	4.1	727	1	BRC1 DROME
33	150.5	4.1	2441	1	CBP_MOUSE

34	150	4.1	1398	1	NCO3 MOUSE
35	149.5	4.1	648	1	GLT0 WHEAT
36	149.5	4.1	1905	1	TAGB DICDI
37	149	4.1	653	1	SF01 MOUSE
38	149	4.1	901	1	Y298 HUMAN
39	148.5	4.0	492	1	YNM4 YEAST
40	148.5	4.0	1365	1	SUZ2 DROME
41	148	4.0	467	1	YP25 CAEEL
42	148	4.0	1128	1	ASP2 HUMAN
43	147	4.0	707	1	FUB2 HUMAN
44	147	4.0	840	1	YA02 HUMAN
45	147	4.0	1157	1	BBC1 YEAST
46	146.5	4.0	639	1	SF01 HUMAN
47	146.5	4.0	1186	1	PKCB HUMAN
48	146.5	4.0	2004	1	MYS3 HUMAN
49	146	4.0	660	1	GLT3 WHEAT
50	146	4.0	1253	1	SHK2 HUMAN
51	146	4.0	1474	1	SHK2 RAT
52	145	4.0	873	1	RX DROME
53	145	4.0	1060	1	YN18 YEAST
54	145	4.0	1230	1	ST20 CANAL
55	144.5	3.9	701	1	CG1 HUMAN
56	144.5	3.9	1177	1	SP97 DICDI
57	144	3.9	418	1	SSXT HUMAN
58	143	3.9	648	1	KAPC DICDI
59	143	3.9	5262	1	MLL2 HUMAN
60	142.5	3.9	2063	1	NCO6 HUMAN
61	142	3.9	467	1	CBPA DICDI
62	142	3.9	861	1	PQ58 CAEEL
63	142	3.9	872	1	SCD5 YEAST
64	142	3.9	1794	1	YDC9 SCHPO
65	142	3.9	5085	1	PCLO RAT
66	141.5	3.9	1374	1	RNC HUMAN
67	141.5	3.9	1638	1	BRW DROME
68	141.5	3.9	2440	1	NCR1 HUMAN
69	141	3.8	478	1	Z207 HUMAN
70	140.5	3.8	843	1	AXN2 HUMAN
71	140.5	3.8	1052	1	CLMN MOUSE
72	140.5	3.8	2442	1	CBP HUMAN
73	140.5	3.8	3669	1	HRX HUMAN
74	140	3.8	375	1	VASP MOUSE
75	140	3.8	700	1	BIB DROME
76	140	3.8	720	1	ABI3 ARATH
77	140	3.8	721	1	FUB2 RAT
78	140	3.8	787	1	FTSH MYCLE
79	140	3.8	1028	1	OVO DROME
80	139.5	3.8	556	1	HIR3 HUMAN
81	139.5	3.8	574	1	MIG1 CANAL
82	139.5	3.8	839	1	GLT5 WHEAT
83	139.5	3.8	3703	1	ABF1 HUMAN
84	139	3.8	794	1	Z148 HUMAN
85	139	3.8	1453	1	NKCR MOUSE
86	138.5	3.8	708	1	VP40 HCMVA
87	138.5	3.8	744	1	REP1 HUMAN
88	138.5	3.8	2039	1	PHF3 HUMAN
89	138.5	3.8	4911	1	MLL3 HUMAN
90	138	3.8	632	1	YKU4 YEAST
91	138	3.8	759	1	MTS1 MOUSE
92	138	3.8	794	1	Z148 RAT
93	138	3.8	1362	1	BRD4 HUMAN
94	137	3.7	1567	1	FMN2 MOUSE
95	136.5	3.7	479	1	PAX3 HUMAN
96	136.5	3.7	838	1	GLT4 WHEAT
97	136.5	3.7	1780	1	YKZ6 CAEEL
98	136	3.7	291	1	GDA2 WHEAT
99	136	3.7	1097	1	CCT DROME
100	135.5	3.7	296	1	PMP3 MOUSE

ALIGNMENTS

RESULT 1

CC	involved in its own circadian regulation. Has no role in regulating circadian clock function in the dark. The activity of the protein may be decreased in long day conditions due to its interaction with phytochrome B (phyB). Can regulate the initiation of flowering independently of phyB. Also involved in responses to nematode parasitism, like the formation of the nematode feeding structure.
CC	-!- SUBUNIT: Interacts specifically with both Pr and Pfr forms of phytochrome B. May form a homodimer.
CC	-!- SUBCELLULAR LOCATION: Nuclear.
CC	-!- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Name=1;
CC	IsolId=082804-1; Sequence=Displayed;
CC	Name=2;
CC	IsolId=082804-2; Sequence=VSP_004042, VSP_004043;
CC	Notes=No experimental confirmation available;
CC	-!- INDUCTION: Expressed with a circadian rhythm showing a peak 14 to 16 hours after sunrise regardless of daylength. Induced in roots after infection by nematodes. Upregulated by auxin and cytokinin and downregulated by abscisic acid and temperature stress.
CC	-!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts in positions 437, 472 and 485.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; Y11994; CAA72719.1; ALT FRAME.
DR	EMBL; AC004747; AAC31242.1; -
DR	EMBL; AC005395; AAM15042.1; -
DR	EMBL; AY136385; AAM97051.1; -
DR	EMBL; BT000185; AAM15504.1; -
DR	PIR; T02630; T02630.
DR	TRANSFAC; T05380; -
DR	Nuclear protein, Transcription regulation;
KW	Phytochrome signaling pathway; Alternative splicing.
KW	DOMAIN 544 585 GLN-RICH.
FT	DOMAIN 442 539 PRO-RICH.
FT	DOMAIN 636 652 THR-RICH.
FT	VARSPPLIC 339 339
FT	VARSPPLIC 340 695 Missing (in isoform 2).
FT	VARSPPLIC 340 695 /FTid=VSP_004043
FT	VARIANT 544 544 Q -> QQQQQQQQQQ (IN CV. WASSILEWSKIJIA).
FT	MUTAGEN 66 73 MISSING: IN ELF3-7; CAUSES EARLY FLOWERING AND LONG HYPOCOTYL PHENOTYPES.
FT	CONFLICT 55 55 M -> R (IN REF. 1).
FT	CONFLICT 196 196 A -> R (IN REF. 1).
FT	CONFLICT 618 618 R -> G (IN REF. 1).
FT	CONFLICT 670 670 K -> KVPFHNAK (IN REF. 1).
SQ	SEQUENCE 695 AA; 77206 MW; 607A0720ED381C08 CRC64;
Query Match 100.0%; Score 3669; DB 1; Length 695;	
Best Local Similarity 100.0%; Pred. No. 7.6e-188;	
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MKRGKDEEKILEPMPFRLHVNDAKGGPRAPRNKMALEYQLSPQFGDHGTNNSRN 60
DB	1 MKRGKDEEKILEPMPFRLHVNDAKGGPRAPRNKMALEYQLSPQFGDHGTNNSRN 60
QY	61 NTSTLVHPGPSQPCGVERNLSVQHLDSAAQATEKFSQMSFNVRSQAQHDQRKMV 120
DB	61 NTSTLVHPGPSQPCGVERNLSVQHLDSAAQATEKFSQMSFNVRSQAQHDQRKMV 120
QY	121 REEDFAVPVYINRSRSGHGRTKSGIEKEKHTPMVAPSSHHHSIRFQEVNQTGSKQNVCL 180
DB	121 REEDFAVPVYINRSRSGHGRTKSGIEKEKHTPMVAPSSHHHSIRFQEVNQTGSKQNVCL 180

GN	NAB3 OR YPL190C.	Best Local Similarity 20.2%; Pred. No. 0.0028;			
OS	Saccharomyces cerevisiae (Baker's Yeast).	Matches 128; Conservative 73; Mismatches 216; Indels 217; Gaps 29;			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RP	Wilson S.M., Oberdorf A.M., Datar K.V., Swedlow J.R., Paddy M.R.,				
RA	Swanson M.S.;				
RA	Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	STRAIN=S288C / AB972;				
RC	MEDLINE=97313271; PubMed=9169875;				
RX	Bussey H., Storms R.K., Ahmed A., Barrell B.G., Badcock K., Benes V.,				
RA	Aratjo R., Aparicio A., Bruckner M., Carpenter J., Cherry J.M.,				
RA	Botstein D., Bowman S., Coster F., Davis K., Davis R.W.,				
RA	Chung E., Churcher C.M., Dipacio T., Dubois E., Dueterhoeft A.,				
RA	Dietrich F.S., Delius H., Friesen J.D., Fritz C., Goffeau A.,				
RA	Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,				
RA	Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,				
RA	Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,				
RA	Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,				
RA	Marathe R., Messenguy F., Mewes H.-W., Mitterpat S., Moestl D.,				
RA	Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,				
RA	Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,				
RA	Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,				
RA	Urestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,				
RA	Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,				
RA	Zhong W.W., Zollner A., Vo D.H., Hani J.;				
RT	"The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.;"				
RL	Nature 387:103-105(1997).				
CC	-1- SUBCELLULAR LOCATION: Nuclear.				
CC	-2- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; U05314; AAA1910.1; -				
DR	EMBL; 273546; CAA97903.1; -				
DR	PIR; S48529; S48529.				
DR	GeneOnline; 144172; -				
DR	SGD; S0006111; NAB3.				
DR	GO; GO:0005654; C:nucleoplasm; IDA.				
DR	GO; GO:0008143; F:poly(A) binding; IDA.				
DR	GO; GO:0006357; P:regulation of transcription from Pol II pro. .; IGI.				
DR	InterPro; IPR000504; RNA_rec_mot.				
DR	Efam; PF00076; rrm; 1.				
DR	SMART; SM00360; RRM; 1.				
DR	PROSITE; PS0102; RRM; 1.				
DR	PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.				
KW	RNA-binding; Nuclear protein.				
FT	DOMAIN 330 401				
FT	DOMAIN 58 71				
FT	DOMAIN 87 93				
FT	DOMAIN 101 106				
FT	DOMAIN 108 115				
FT	DOMAIN 116 127				
FT	DOMAIN 128 137				
FT	DOMAIN 603 608				
FT	DOMAIN 644 648				
FT	DOMAIN 698 703				
FT	DOMAIN 723 728				
FT	DOMAIN 765 768				
FT	DOMAIN 769 784				
FT	DOMAIN 802 AA; 90438 MW; CD6D2C7F24A44993 CRC64;				
CC	-----				
CC	Query Match				
CC	5.3%; Score 193.5; DB 1; Length 802;				

FT	DOMAIN	231	244	CRIB.	
FT	DOMAIN	680	935	PROTEIN KINASE.	
FT	DOMAIN	16	19	POLY-PRO.	
FT	DOMAIN	210	213	POLY-SER.	
FT	DOMAIN	311	315	POLY-ASN.	
FT	DOMAIN	384	389	POLY-ASN.	
FT	DOMAIN	392	395	POLY-ASN.	
FT	DOMAIN	418	421	POLY-PRO.	
FT	DOMAIN	453	462	POLY-GLN.	
FT	DOMAIN	468	476	POLY-GLN.	
FT	DOMAIN	572	576	POLY-GLN.	
FT	NP BIND	686	694	ATP (BY SIMILARITY).	
FT	BINDING	710	710	ATP (BY SIMILARITY).	
FT	ACT SITE	803	803	BY SIMILARITY.	
FT	SEQUENCE	971 AA;	196889 MW;	ADGF0DBBC6CF624B	CR064;

```

Query Match      4.9%  Score 179; DB 1; Length 971;
Best Local Similarity 25.0%  Pred No. 0.02;
Matches          93; Conservative

```

QY	377	KVSAKSYPVVKLLPSEFLVKKPPVPHVVVVKQGRGSEKTDQHKMSSAENVVGRLSNQGHQ	436
DB	337	KPPAKS-TVQCFKPSRAAPKPTPYHLTOLNGSS-----HQTSSS-----GSPSSGNNN	386
QY	437	QSNVMPFANNPPASPAN-----GYCFP-----POPPSG-----NH-----QOWL	472
DB	387	NNNSTNNNTKNVSLNLMKSELIPARAPPPPTSGTSDTYSNKNHQDRSGYEORQ	446
QY	473	IPVMSPSSEGLIYKPH-----PGMAHTGHYGGYGHYMPMTVMWPQHPGMPG	520
DB	447	QRTDSSQOQOQKQHQYQKSKQOQOQPOOPLSSHOGG-TSH-----IPKQVPPTLPSGG-P	501
QY	521	P--PCNGVFPYPYGNMPTI-----MNPYCSSQOQOQOQNEQMNQFGHPGNLQNTQQOQOR	573
DB	502	PTQAASGKSPSKLHPDLKIQOQTNNYIKSGTIDANQVDGDAKFIKPFNLQSKSKSQOL	561
QY	574	SDNEPAP-----QOQOQPTKSYPRARKRGQSTGSS-----PSGQGISGSKSF	617
DB	562	ASKQSPSPSSQOQOQKPMTSH-----GLMGTSHSVTKPLNPVNDPIKPLNLKSSKSK	613
QY	618	RPFAAVEDSINNPAPECTMTTTTTTTRTTTQTTRDGGGVTRIVKVVPHPNAKLASENAA	677
DB	614	E---ALNETSGVSKTSPSTDNSKPTAPAS-----GPAVTKTAK-----	649
QY	678	RIFQSIQBERKR	689
DB	650	-----OLKKERER	657

RESULT 7	ID	MAM_DROME	STANDARD;	PRT; 1596 AA.
AC	P21519;			
DT	01-MAY-1991	(Rel. 18, Created)		
DT	01-MAY-1991	(Rel. 18, Last sequence update)		
DT	15-MAR-2004	(Rel. 43, Last annotation update)		
DE	Neurogenic protein mastermind.			
GN	MAM.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Canton-S;			
RX	MEDLINE=91065516; PubMed=1701150;			
RA	Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,			
RA	Yeavobnick B;			
RT	"The Drosophila neurogenic locus mastermind encodes a nuclear protein			
RT	unusually rich in amino acid homopolymers.";			
RL	Genes Dev. 4:1688-9100(1990).			
CC	!- FUNCTION: May have a regulatory function possibly in association			

```

QY 528 ---PPYGM-----MPTIMP-----YCSSQQQQQQQQPNQSMNQFGHPGNLQ 565
Db 792 MPPDHGMKGFPFPHGPFVVRPGMPGLGGPGPGGSEDRDGRQQPQQQQQQ-----QQ 845
QY 566 NTQQQQQRSDNEPAQQ-----QQQPTKSYPRARKSRQGGTSGSPGQPGIGISGSKSPRPFQA 622
Db 846 QQQQQQQQQQQQPPQQSQTOQQPAPSQQPAQAQQPQQQFRNDNRQQQFNSGRDQER-FGR 904
QY 623 VDEDSNNINNAPEQ 635
Db 905 RSFGSRVENDRER 917

RESULT 6
CLAA4_CANAL STANDARD; PRT; 971 AA.
AC O14427;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase CLAA4 (EC 2.7.1.1.-).
GN CLAA4.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
RN NCBI_TaxID=5476;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97411146; PubMed=9259554;
RA Leberer E., Ziegelbauer K., Schmidt A., Marcus D., Dignard D., Ash J.,
RA Johnson L., Thomas D.Y.;
RT "Virulence and hyphal formation of Candida albicans require the
RT Ste20p-like protein kinase CaClap4.";
RL Curr. Biol. 7:539-546(1997).
CC -!- FUNCTION: Essential for virulence and morphological switching
CC (hyphal formation) of C.albicans.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC STE20 subfamily.
CC -!- SIMILARITY: Contains 1 CRIB domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EHSP; U67996; AAB68613.1; -.
DR HSSP; Q63450; 1A06.
DR InterPro; IPR0000095; FAKbox/Rhoindng.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SMO0220; S_TKC; 1.
DR SMART; PS50108; CRIB; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR Transferase; Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 67 178 PH.

```


Neumann S., Argiriou A., Vitale D., Liguri R., Piravandi E., Massenel O., Quigley F., Clabaud G., Muenchlein A., Felber R., Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S., Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A., Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse C., Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek T., Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., Zaccarilli P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L., Sekhon M., Murray J., Sheet P., Cordes M., Abu-Thomas J., Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J., Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W., Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C., Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C., Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R., Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S., Granat S., Shondy N., Hasegawa A., Hameed A., Lodhi M., Johnson A., Chen E., Marra M.A., Martienssen K., McCombie W.R.; "Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana."; Nature 402:769-777(1999).

RT Nature 402:769-777(1999).

RL -!- FUNCTION: Binds the poly(A) tail of mRNA.

CC -!- TISSUE SPECIFICITY: Roots and shoots.

CC -!- MISCELLANEOUS: A.thaliana contains at number of PABP genes which are expressed in an organ-specific manner.

CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; LI9418; AAA61780.1; --

CC EMBL; AL021961; CAAL7561.1; --

CC EMBL; AL161584; CAB80128.1; --

CC PIR; T05425; T05425.

CC HSP; P11940; LCVI.

CC InterPro; IPR002343; Hsd_Sxl_RNA.

CC InterPro; IPR002004; PABP/HSCt.

CC InterPro; IPR006515; PABP_1234.

CC InterPro; IPR000504; RNA_rec_mot.

CC Pfam; PF00658; PABP; 1.

CC Pfam; PF00076; rrm; 4.

CC PRINTS; PR00961; HUDSLRNA.

CC SMART; SMO0517; POLYA; 1.

CC SMART; SMO0360; RRM; 4.

CC TIGRfam; TIGR01628; PABP-1234; 1.

CC PROSITE; PS01012; RRM; 4.

CC PROSITE; PS00030; RRM_RNP_1; 3.

CC RNA-binding; Nuclear protein; Repeat; Multigene family.

CC DOMAIN 36 114 RNA-BINDING (RRM) 1.

CC DOMAIN 124 196 RNA-BINDING (RRM) 2.

CC DOMAIN 215 292 RNA-BINDING (RRM) 3.

CC DOMAIN 318 395 RNA-BINDING (RRM) 4.

CC SEQUENCE 629 AA; 68672 MW; 0F809818D08BDC7E CRC64;

Query Match 4.5%; Score 164.5; DB 1; Length 629;

Best Local Similarity 20.7%; Pred. No. 0.072;

Matches 126; Conservative 71; Mismatches 210; Indels 203; Gaps 30;

QY 67 HPGSGPCGVERNLVQHLDS-----AAQATEKFSQMSFMENESSAQHDKVRE 122

DB 113 HRDSVRESGA-GNIFKNLDESIDKALHDTFSFGNIVSCKVAVDSGG----- 162

QY 123 EEDFAVPVYINRSQSQRKTSGLIEKHTPMVAPSSHHHSIRFQEVNQTSGKNVCLAT 182

[illegible]

Thu Jul 29 08:52:25 2004

Best Local Similarity 20.5%; Pred. No. 0.096;
Matches 118; Conservative 68; Mismatches 204; Indels 185; Gaps 25;

QY 176 QNVCLATCKSEV-----RDQKANA-----RSGGFVISLDVSV 209
Db 128 KNSCLRTQKQKQVFWFNVPHDKLMADALERDLKKEKMGQRTTWAHREPALSFDYDESS 187
QY 210 TBEIDLEKSASHDRVNDYNASLROESNRRLYDGGKTRLDKTDNGASHLATENH----- 265
Db 198 SLYTQLGKHEMTEQKINDAATSTSTNTATTLTDGVSSGLNNTTSGGSDSATSTHNNNE 247
QY 266 --SQEGHSGPEDIDNDREY-SKSRACASLQOLNE-----EASDDVSDSDMSVDSISIDV 316
Db 248 ASGTSPNSGSEK---SSPEYTTTARGRDEFGFLNEATPQSKANSYDEDDFLDYINQTTQ 304
QY 317 SPDDVVGILGQKFEWFEARAKAIAQORVAVOLFELHRLIKVKQLIAAASPDLLDLSFL- 375
Db 305 NSEDIYITL-----DANYQAG-----SYANMIEDNYSFLDALTFLIP 340
QY 376 -----GKVSAKSYPVKKLLPSEFL-----VKPPLPHVVVKQGDSEKTDQHKWESSA 422
Db 341 PSLGVPVTGTAATATTGNQVAFNDXYLLQOAPITRPLPT-----SS 382
QY 423 ENVVGRK-----SNQGHQOSNMPFANPPASAPNGYCFPPQPPPSGNH-QQWL 472
Db 383 STISGLLQPKSAAKFFSLOSANGGEEFFPAYQNDP-STANAGFV-----PPISAKYATQFA 437
QY 473 IPWMSFSEGLIYKPHPGMAHTGHYGG-----YUGH-----YMPPT--PMVMPQYHP----- 515
Db 438 TROVATPTYIKALPQTGAATAATNGGQPOQYDQATGNFPAEIPVLYNVVHPESEYWT 497
QY 516 -----GNGFPPPGNGYF-----PPY-----GMPTTINPVCSS 543
Db 498 NNSGAVATTAATAAPMYDASGFPPIQNSYMWNEHWPYQYMNNSGAMIGMIPPH--- 554
QY 544 QQQQQQQPQNMQFGHPGMLNLTQQQQQSDNEPAPQOQQOQPKSYPRARKSRQSTGS 603
Db 555 QQQQQQQQ-----QQTAMGQSMRLRQOQQOQQO-----QQQQQPSSTWTKKKKQIHSENN 604
QY 604 SPSPQOQISGSKSPRPFPAAVDEDSNINNAPQWT 638
Db 605 KSLSSKG-----GGITKKSHDNNHKS VKT 629

RESULT 17
ATX1 HUMAN
ID ATX1 HUMAN STANDARD; PRT; 816 AA.
AC P54253; Q9UJG2; Q9Y4J1
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ataxin-1 (Spinocerebellar ataxia type 1 protein).
GN SCAL OR ATX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Cerebellum;
RX MEDLINE=95038838; PubMed=7951322;
RA Banfi S., Servadio A., Chung M.-Y., Kwiatkowski T.J. Jr., McCall A.E.,
RA Duvick L.A., Shen Y., Roth E.J., Orr H.T., Zoghbi H.Y.;
RT "Identification and characterization of the gene causing type 1
RT spinocerebellar ataxia."
RL Nat. Genet. 7:513-519(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Tubby B.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 189-231 FROM N.A.
RX MEDLINE=96177682; PubMed=8634720;

RA Quan F., Janas J., Popovich B.W.;
RT "A novel CAG repeat configuration in the SCAL gene: implications for
RL the molecular diagnostics of spinocerebellar ataxia type 1.";
RN Hum. Mol. Genet. 4:2411-2413(1995).
[4]
RP SELF-ASSOCIATION SITE.
RX MEDLINE=97252384; PubMed=9097953;
RA Burright E.N., Davidson J.D., Duvick L.A., Koshy B., Zoghbi H.Y.,
RA Orr H.T.;
RT "Identification of a self-association region within the SCAL gene
RT product, ataxin-1.";
RL Hum. Mol. Genet. 6:513-518(1997).
[5]
RN RNA-BINDING DOMAIN.
RX MEDLINE=21065637; PubMed=11136710;
RA Yue S., Serra H.G., Zoghbi H.Y., Orr H.T.;
RT "The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding
RT activity that is inversely affected by the length of its polyglutamine
RT tract.";
RL Hum. Mol. Genet. 10:25-30(2001).
CC -!- FUNCTION: Binds RNA in vitro. May be involved in RNA metabolism.
CC The expansion of the polyglutamine tract may alter this function.
CC -!- SUBUNIT: Interacts with LAMP and ALU.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=At least 2 isoforms are produced;
CC Name=1;
CC IsoId=P54253-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Widely expressed throughout the body.
CC -!- POLYMORPHISM: The poly-Gln region of SCAL is highly polymorphic (4
CC to 39 repeats) in the normal population and is expanded to about
CC 40-83 repeats in scal patients. Longer expansions result in
CC earlier onset and more severe clinical manifestations of the
CC disease.
CC -!- DISEASE: Defects in SCAL are the cause of spinocerebellar ataxia
CC type 1 (SCAL) [MIM:164400]; also known as olivopontocerebellar
CC atrophy 1 (OPCA 1). SCAL is an autosomal dominant
CC neurodegenerative disorder characterized by progressive neuronal
CC loss in the cerebellum, brain stem and spinocerebellar tracts.
CC Clinical features are cerebellar ataxia, dysarthria,
CC oculopalpebral paresis, muscle wasting and neuropathy. Onset of the
CC disease usually occurs in the third or fourth decade of life and
CC death occurs ten to twenty years later.
CC -!- MISCELLANEOUS: The self-association seems to be necessary to form
CC nuclear aggregates.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79204; CAA55793.1; -;
CC EMBL; AL009031; CAA15622.1; -;
CC EMBL; S82497; AAD14401.1; -;
CC EMBL; S46268; S46268.
CC EMBL; HGNC:10548; SCAL.
CC MIM; 601556; -;
CC MIM; 164400; -;
CC GO; GO:0005737; C:cytoplasm; TAS.
CC GO; GO:0005634; C:nucleus; TAS.
CC InterPro; IPR003652; Ataxin.
CC SMART; SM00536; AXH; 1;
CC RNA-binding; Nuclear protein; Polymorphism; Triplet repeat expansion;
CC RNA-Alternative splicing.
CC FT DOMAIN 197 226
CC FT DOMAIN 495 605
CC FT DOMAIN 541 767
CC FT DOMAIN 795 798
CC POLY-GLN.
CC SELF-ASSOCIATION SITE.
CC RNA BINDING.
CC NUCLEAR LOCALIZATION SIGNAL
CC (BY SIMILARITY).


```

CC CC Name=2;
CC CC IsoId=09UGU0-2; Sequence=VSP_003984, VSP_003985;
CC CC TISSUE SPECIFICITY: Expressed in most tissues, except in ovary and
CC CC prostate. Isoform 1 is exclusively expressed in brain, heart and
CC CC testis, and this form predominates in liver and kidney. Isoform 2
CC CC predominates in lung.
CC CC -!- DOMAIN: The atypical PHD domain functions as a negative modulator
CC CC of cofactor binding (By similarity).
CC CC -!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
CC CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC CC -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC CC frameshift in position 1932.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AY007595; CAB28930.1; -
CC CC EMBL; AL031346; CAB42440.1; ALT INIT.
CC CC EMBL; AL021878; -; NOT ANNOTATED_CDS.
CC CC EMBL; AB006630; BAA22961.1; -
CC CC EMBL; U19345; AAC36392.1; ALT_FRAME.
CC CC Genew; HGNC:11631; TCF20.
CC CC MIM; 603107; -
CC CC GO; GO:0005634; C:nucleus; NAS.
CC CC GO; GO:0003677; F:DNA binding; NAS.
CC CC GO; GO:0003713; F:transcription co-activator activity; NAS.
CC CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
CC CC InterPro; IPR001945; Znf_PHD.
CC CC SMART; SM00249; PHD; 1.
CC CC KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
CC CC KW Nuclear protein; Alternative splicing.
CC CC FT DOMAIN 1170 1191 LEUCINE-ZIPPER.
CC CC FT DOMAIN 1254 1268 NUCLEAR LOCALIZATION SIGNAL
CC CC (BY SIMILARITY).
CC CC FT DNA_BIND 1537 1551 A.T HOOK.
CC CC FT DOMAIN 1576 1600 NUCLEAR LOCALIZATION SIGNAL
CC CC (BY SIMILARITY).
CC CC FT DOMAIN 1785 1792 NUCLEAR LOCALIZATION SIGNAL
CC CC (BY SIMILARITY).
CC CC FT ZN_FING 1884 1935 PHD-TYPE (ATYPICAL).
CC CC FT DOMAIN 62 70 POLY-ALA.
CC CC FT DOMAIN 174 182 POLY-GLN.
CC CC FT DOMAIN 203 262 SER-RICH.
CC CC FT DOMAIN 310 322 POLY-GLN.
CC CC FT DOMAIN 1556 1564 POLY-PRO.
CC CC FT VARSPLIC 1934 1938 PEPIC -> VRLWR (in isoform 2).
CC CC FT VARSPLIC 1939 1960 Missing (in isoform 2).
CC CC FT CONFLICT 122 122 /FTId:VSP_003985.
CC CC FT CONFLICT 200 200 R -> Q (IN REF. 2).
CC CC FT CONFLICT 200 200 K -> Q (IN REF. 2).
CC CC FT SEQUENCE 1960 AA; 211798 MW; F459146C05B6D2A CRC64;

Query Match 4.4%; Score 161; DB 1; Length 1960;
Best Local Similarity 19.1%; Pred. No. 0.43;
Matches 147; Conservative 90; Mismatches 261; Indels 270; Gaps 32;

QY 17 RLHVNDADKGGPRAPRNKALYEQLSIPSRFGHGTGHTNRSNNITSLVHPGPSQP-- 74
DB 1015 KLMSPGRSGRGGDPHH-----MNPHTFSERANRSLHTPSPNSETLA 1060
QY 75 -----CGVERNLVQHLDSANQATEKFSQMS-FMENVRSSAQHDQ- 116
DB 1061 SAYHANTRAHVGDNAGNSQI---HYKRMYYQQQPEEYKDWSSGSAQGVIAAAQHQE 1117
QY 117 --RKVREEDFAVPVYINRSRSGHGRKSGIEKEKHTPMVAPS----- 159
DB 1118 GPRK-----SPRQQQLFDRVRSPLKNDKGMVGPVGYTHDPSAQEAGRC 1163

```

```

QY 160 -----SHSIRFOEVNQTSKQNVCLATCSKPEVRDQ----- 191
DB 1164 LMSDGLPNKGMELKHGSKLQSCWDLRSQTSPAKSSGPPGMSSQKRYGPPHETDGHGL 1223
QY 192 VKANARSGGFVILSDVSVTTEEIDLEKSASSHDRVNDYNASL-----RQSRNRLYRDGGK 246
DB 1224 ABATQSSKPGSVMLRLPGQEDHSSQNPLIMRRVRVSIFIPSKRQSDVKNSTEDKGR 1283
QY 247 TRLKDTDNGAE-----SHLATNHSQEGHSGPE-DIDNDREYSKSRACASLQQINBEAS 299
DB 1284 L-LHSSKEGADKAFNSVAHL-----SHSQDIKSIKPRDSSKDLPSDPCNCFV----- 1331
QY 300 DVVSDSDSDVSDISIDVSPDDVVGILGQKRFWRKAKAIANQORFAVLQELHRLIKVQK 359
DB 1332 -----TITSPAKTKILP-----PRKGRGLKLEAIVQKITSP---NIRRSASSNS 1372
QY 360 LIAASPDLLLEISFL-----GKVSAKSYPPVKK-----LLPS--EFLVKPPLPH 401
DB 1373 ABAGGDTVTLLDILSLKSGPPEGGSVAVQADADIEKRKGEVASDLVSPANQELHVEKPLR 1432
QY 402 VVYKQRGDSKTDQHKWESSAENVV-----GRL-----SNOGHQOS-----NY 440
DB 1433 SSEEWGSGVD--DKVKTETHAETVTAGKEPPGAMTSTTSQKPSGNQGRPDGSLGTAPLI 1490
QY 441 MPFANNPPASAPNGYCFPPQPPPSGNHQWLIPVMSPEGLLYKHPGMAHTGHYGGY 500
DB 1491 FPDCKNVP-----PVGILAFENPKAEKENDVTITSPKQ----- 1526
QY 501 GHYMPTEMVMPQYHGMGFFPGNGYFP-----EYGMPTTINPFCSSQQQQQQOPNE 553
DB 1527 -----GFPF--KGYPSPKKGRIQSV-----NKQKQQQP-- 1556
QY 554 QMNQFHPGNLQNTQQQQRSNDNEPAPQQQQQPTK-----SVPRAKSRQSGTSGSPSGP 608
DB 1557 -----PPPPPPPPQIPEGSSADGEPKPKQRORRRKPKCAOPKPKTKQAVPIVEPOEP 1610
QY 609 Q-----GLSGSKSRFPAVDEDSNIN-----NAPEQMT 638
DB 1611 EIKLYATQPLDKTDANKNSFYPIHVKNKCELGAVCTIINAEEDQT 1658

```

RESULT 20

```

Y553 HUMAN
ID Y553 HUMAN STANDARD; PRT; 1089 AA.
AC Q9UKJ3; O60300;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0553.
GN KIAA0553.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=96288581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99408744; PubMed=10477733;
RA Thornton M.A., Poncz M., Korostishevsky M., Yakobson E., Usher S.,
RA Seligson U., Peretz H.;
RT "The human platelet alphaIIb gene is not closely linked to its
RT integrin partner beta3.";
RL Blood 94:2039-2047(1999).

```

548 QQQNEQNMQFGHPGNLQNTQQQQQSDNEPAPQOQQOQFTKSPARKSROGSTGSSPSG 607
888 SLAPLESQPIFTPEEMEKYSKIQAAQOQQIQQQLLAKQVKAFPP-----ASAALAPAT 940
608 PQGISGSKSRFPFAAVEDSNINNAPEQTMTTTTTTTTTTTTTRDGGGVTTRVKKVVP 667
941 P-----ALQPIHI-----QQPATASATSTTVQHAILQ--HHAASAAAAGIHPH 983

RESULT 21
CBP1 CAEEL STANDARD; PRT; 2056 AA.
ID CBP1 CAEEL
AC P34545;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein cbp-1.
GN CBP-1 OR R10E11.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton L., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fultons M., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden K.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38 (1994).
[2]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=b;
CC IsoId=P34545-1; Sequence=Displayed;
CC Name=a;
CC IsoId=P34545-2; Sequence=VSP_000557;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 Z2-type zinc finger.
CC -!- SIMILARITY: Contains 2 TAZ-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z29095; CAA82353.2; -;
CC EMBL; Z29095; CAD18875.1; -;
CC PIR; G88564; G88564.
CC WormPep; R10E11.1a; CE28069.
CC WormPep; R10E11.1b; CE21117.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR003101; KIX.
CC InterPro; IPR000197; TAZ finger.
CC InterPro; IPR001965; Znf PHD.
CC InterPro; IPR000433; Znf_ZZ.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AB011125; BAA25479.1; ALT_INIT.
CC EMBL; AF160252; AAF03681.1; -;
KW Hypothetical protein.
FT DOMAIN 263 266 POLY-LYS.
FT DOMAIN 332 336 POLY-ARG.
FT DOMAIN 359 362 POLY-GLY.
FT DOMAIN 425 430 POLY-GLU.
FT DOMAIN 456 461 POLY-SER.
FT DOMAIN 971 977 POLY-ALA.
FT DOMAIN 1059 1065 POLY-ALA.
FT CONFLICT 207 207 K -> E (IN REF. 1).
FT CONFLICT 267 267 E -> K (IN REF. 1).
FT CONFLICT 278 278 P -> H (IN REF. 1).
FT CONFLICT 846 846 G -> S (IN REF. 1).
SQ SEQUENCE 1089 AA; 117999 MW; 4EF687F9D81A16A5 CRC64;

Query Match 4.3%; Score 158; DB 1; Length 1089;
Best Local Similarity 18.7%; Pred. No. 0.31;
Matches 146; Conservative 115; Mismatches 309; Indels 210; Gaps 36;

QY 2 KRQKDEKILEMPRLHVNDKDG-----GPRAPRNKMAV---EQLSIPSRFG 50
DB 300 KRKKKKNSAPA-----DSERGPKPEPGSGPAPPRRRRAODSRRSLPAEE-G 351
QY 51 DHGTMSRSNNTSLVHPG-----PSS--QPCGVERNLSVQH-----LDSSA 90
DB 352 SSGKXDEGGGGSSQDHGGRKHKGELPPSCQRRAGTKRSRSHRSQPSGGDESDAS 411
QY 91 ANQATEKVFQMGFME-----NVRSAQHDPQKVR----- 121
DB 412 SHRLHQSPQYSEEEEDSGSHSRSRSGRRHSHSRSSSSSSDASSDQSCYS 471
QY 122 EEDFAVPVYIN--SRQSQHGRTKSGTEKEKHTPWAPSSHSTRFQVNTGSKQNVCL 180
DB 472 RQSYSDSDSYDSDRGRHRSKRHSDSDSY-----ASSKRSKRHYSSDDDDYSLSC- 526
QY 181 ATCSKPEVRQVKANARSGGFVILDVSVTEIDLEKSASGH--DRVNDYNASLRQPSRN 238
DB 527 -SQRSRSRSHTRRSRSGRSRSCSRSSCRSRKRSRSTTAHSWQSRKSYSDRSRSTRS 585
QY 239 RLYRDGGKTRUKTDNGAESHLATEN-----HSQ-----EGHSGPEDIDNDRE 281
DB 586 PSQSGSKRKRSGWGHSPERHSRRDFTRSKYRSQSPHYFRSGRGEQPGKKDGRGD-- 643
QY 282 YKSRACASLOQINEEASDDVDSDWDS--LSSIDVSPDWDVGLGQKRFWRARKALAN 339
DB 644 -----DSKATGPPSQNSNITGSGEGDCSPED-----KNSVTA 677
QY 340 QQRVFAVOLFFELHRLIKVKQLIAASPDLLDLBEISFLGKVSASKYPVKLLSEFLVKPPL 399
DB 678 KILLEKIOSRKVERKPSVSEEQATPNKAGPKL---KDPQGYGPKLTPPS-LGNKPV 732
QY 400 PHVVVQRGDSEKTDQHKMSESAENVVGRNLNOGHQHSNYPFPANNRP-ASPAFNGYCF 458
DB 733 P--LIGKLPATRKPNKKBESGLE-----RGEEQES--ETBEGPPGSSDALFGHQF 780
QY 459 PPO-----PPP-----SGNHQQLVLPVMSFSEGLIYKPHPGMAHTGHYGYTGH 502
DB 781 PSEETGTPLDDPPPEESKSGEATADHP--VAELGTPTAHSDCYCPGPTTISH-----N 829
QY 503 YNPTPW--VMPQYHFGMGFPP-----PCNGYFPPYGMPTMMPYCS-----QQQQ 547
DB 830 YLPDPDSDGDTLESLDGSGQPGFVSSLLPIAPDLEHFTSYA--PPSGDFSIESTDGAEDA 887

```

DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02172; KIX; 1.
DR Pfam; PF02135; zf-TAZ; 2.
DR Pfam; PF00569; zf; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00551; ZnF_TAZ; 2.
DR SMART; SM00291; ZnF_ZZ; 1.
DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR PROSITE; PS50014; BROMODOMAIN 2; 1.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PS01134; ZF_TAZ; 2.
DR PROSITE; PS01357; ZF_ZZ 1; 1.
DR PROSITE; PS01355; ZF_ZZ 2; 1.
KW Bromodomain; Metal-binding; Zinc; Zinc-finger; Alternative splicing;
Repeat.
FT ZN FING 399 505 TAZ-TYPE 1.
FT DOMAIN 881 953 BROMODOMAIN.
FT ZN FING 1493 1534 ZZ-TYPE.
FT ZN FING 1550 1631 TAZ-TYPE 2.
FT DOMAIN 1687 2008 GLY/GLN-RICH.
FT VARSPPLIC 467 478 SDTQTQTKGVSF -> F (in isoform a).
FT FTID=VSP 000557
SQ SEQUENCE 2056 AA; 227179 MW; 949FF4608C634F01 CRC64;

Query Match 4.3%; Score 158; DB 1; Length 2056;
Best Local Similarity 24.9%; Pred. No. 0.65;
Matches 75; Conservative 29; Mismatches 117; Indels 80; Gaps 16;

QY 439 NYMPFANNPPA-----SPAPNGCYCFPPQPPPSQHQQWLIPVMSPE----- 480
Db 136 NGTPNMSPPSGRVDPSPGG-----PQPPGPGQ-----PQNRPGQPGMFGDQQQQMM 185
QY 481 -GLIYKPHGMAHTGHYGYGHYMPTEPMVPOYHPGM--GFPFPG-----NGYFP 528
Db 186 MGAQGOQFGMHRHYEYA--QGPPPGAQGMPOGYFSGVSGGTTPQPMGRGMWNGAMP 243
QY 529 FYGMPT-----IMNPGCSQQ-----QQQQQPNQOMNFGHPGLNLTQQ-Q 570
Db 244 RSGPMPTQGRGIPNPQQANQMFMPTDRQFMHGQYQGRPEFMQYGRFGYPMWHQGM 303
QY 571 QQRSDNEP--APQQQQQPTKSVPRARKSQSGTSSFS---GPGTSGSKSFRPFAVDE 625
Db 304 MDSNGQPLRGPNQMMNSNGHP-----GMSHGPPNGQPGPQAAAAQAQAQAQAQA 356
QY 626 DSNINNAPEQTMTTTRTTTTRTTTQTDRDGGVTRVIVKVPVHNKLASENAARIFQSIOE 685
Db 357 QAQA--AAQQQQQQQREQAQAQA--RNGAGRA-----TTPGSSMLATHQDPEKRLIQ 408
QY 686 E 686
Db 409 Q 409

RESULT 22
PROS_DROME
ID PROS_DROME STANDARD; PRT; 1403 AA.
AC F29617; Q95SP0; Q9UGA2; Q9VGP8;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein prospero.
PROS OR CGI7228.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92069760; PubMed=1720353;
RA Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
RT "Prospero is expressed in neuronal precursors and encodes a nuclear
protein that is involved in the control of axonal outgrowth in
Drosophila.";
Cell 67:941-953(1991).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92171948; PubMed=1540176;
RA Matsuzaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;
RT "Cloning of the Drosophila prospero gene and its expression in
ganglion mother cells.";
Biochem. Biophys. Res. Commun. 182:1326-1332(1992).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93083413; PubMed=1842358;
RA Chu-Lagraff O., Wright D.M., McNeil L.K., Doe C.O.;
RT "The prospero gene encodes a divergent homeodomain protein that
controls neuronal identity in Drosophila.";
Development Suppl. 2:79-85(1991).
[4]
RP SEQUENCE FROM N.A. (ISOFORM C).
RX MEDLINE=20503846; PubMed=11051550;
RA Xu C., Kauffmann R.C., Zhang J., Kladny S., Carthew R.W.;
RT "Overlapping activators and repressors delimit transcriptional
response to receptor tyrosine kinase signals in the Drosophila eye.";
Cell 103:87-97(2000).
[5]
RP SEQUENCE FROM N.A.
STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.G.,
RA Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu K., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
[6]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,

```

Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 Lewis S.E.;
 "Annotation of the Drosophila melanogaster euchromatic genome: a
 systematic review";
 Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 [7]
 SEQUENCE FROM N.A. (ISOFORM A).
 STRAIN=Berkley; TISSUE=Head;
 MEDLINE=22426066; PubMed=12537569;
 Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 George R.A., Guarini H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 Rubin G.M., Celnik S.E.;
 "A Drosophila full-length cDNA resource";
 Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 [8]
 SIMILARITY TO C.ELEGANS CEH-26.
 MEDLINE=94212446; PubMed=7909177;
 Buerklin T.R.;
 "A Caenorhabditis elegans prospero homologue defines a novel domain";
 Trends Biochem. Sci. 19:70-71(1994).
 -!- FUNCTION: Required for proper neuronal differentiation of most or
 all neurons and their precursors in central and peripheral nervous
 systems, axonal outgrowth and pathfinding. Not required for the
 specification of neuronal identity. May regulate transcription by
 binding to DNA.
 -!- SUBCELLULAR LOCATION: Nuclear.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=4;
 Comment=Experimental confirmation may be lacking for some
 isoforms;
 Name=C; Synonyms=L;
 IsoId=P29617-1; Sequence=displayed;
 Name=A;
 IsoId=P29617-2; Sequence=VSP_002307, VSP_002308;
 Name=B;
 IsoId=P29617-3; Sequence=VSP_002307, VSP_002309;
 Name=S; Synonyms=D;
 IsoId=P29617-4; Sequence=VSP_002308;
 -!- TISSUE SPECIFICITY: Neuronal precursors. Expressed in the
 developing CNS, lens-secreting cone cells of the eye, and midgut.
 -!- DEVELOPMENTAL STAGE: Expressed in neuronal precursors early during
 formation.
 -!- SIMILARITY: Belongs to the Prospero homeobox family.
 -!- SIMILARITY: Contains 1 homeobox domain.
 -!- CAUTION: Ref.7 sequence differs from that shown due to a
 frameshift in position 1122.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M81389; AAA28841.1; -;
 EMBL; D10609; BAA01464.1; -;
 EMBL; Z11743; CAA77802.1; -;
 EMBL; AF190403; AAF05703.1; -;
 EMBL; AE003691; AAF54628.2; -;
 EMBL; AE003691; AAN13500.2; -;
 EMBL; AE003691; AAN13501.2; -;
 EMBL; AY060680; AAL28228.1; ALT_FRAME.
 PIR; S24548; S24548.
 PDB; 1MJJ; 04-DEC-02.
 Flybase; Fgn0004595; pros.
 GO; GO:0045179; C:apical cortex; IDA.
 GO; GO:0045180; C:basal cortex; IDA.
 GO; GO:0005634; C:nucleus; IDA.
 GO; GO:0003700; F:transcription factor activity; NAS.
 GO; GO:0007409; P:axogenesis; IMP.
 GO; GO:0007417; P:central nervous system development; IMP.

DR GO:0016358; P:dendrite morphogenesis; IMP.
 DR GO:0007422; P:peripheral nervous system development; IMP.
 DR GO:0045664; P:regulation of neuron differentiation; IMP.
 DR InterPro; IPR007738; Prox1.
 DR Pfam; PF05044; Prox1; 1.
 KW Nuclear protein; Transcription regulation; DNA-binding; Homeobox;
 KW Developmental protein; Alternative splicing; 3D-structure.
 FT DOMAIN 4 12
 FT POLY-SER. 28 31
 FT POLY-ASN. 32 35
 FT POLY-ALA. 188 191
 FT GLN-RICH. 216 264
 FT ASN-RICH. 270 286
 FT SER-RICH. 318 354
 FT POLY-ASP. 431 437
 FT POLY-ALA. 505 508
 FT DOMAIN 700 1048
 FT DOMAIN 934 937
 FT DOMAIN 991 998
 FT DOMAIN 1074 1082
 FT HIS-RICH. 1137 1137
 FT HOMEBOX (ATYPICAL). 1241 1303
 FT DNA_BIND 1304 1403
 FT DOMAIN 916 916
 FT VARSPLIC
 FT
 FT VARSPLIC 1216 1244
 FT
 FT VARSPLIC 1236 1403
 FT
 FT CONFLICT 76 98
 FT
 FT CONFLICT 120 144
 FT
 FT
 Query Match 4.3%; Score 156; DB 1; Length 1403;
 Best Local Similarity 17.8%; Pred.No. 0.53;
 Matches 155; Conservative 138; Mismatches 336; Indels 240; Gaps 36;
 Qy 4 GKDEKILEPMFPR--LHVNDADKGGPRAPPRKMA-----LYEQLSIPSRFG 50
 Db 72 GSSSAKMLNLFGRMKQAQDATSGLPQSLDNAMLAAMETATSAELIISLSTSKLLQ 131
 Qy 51 DHGTMS-----RSNTSTIVHPGSS-----OPCG-----VERNLSVOHL 86
 Db 132 QQNNNSIAPNSTPMNGTNASISPGSAHSSSHSHQGVSPKGSRRYSACSDRSLEAAA 191
 Qy 87 DSSAANQATEKFVSQM----SFMENVRSSAQHD-----ORKMVBREDFAVP 129
 Db 192 DVAGGSPRAASVSSSLNGGASSGEGHQSLQHLVAHMLRNLIQKELMQLDELRTA 251
 Qy 130 VYINRSRSHGRYTGIEKHTPMVA-PSSHSIRFQEVNQTGSKNVCIATCKPEV 188
 Db 252 MQQQQQQLKEKEQLHSLKLNNNNNNIAATANNNTTWESINLIDDEMAKIKSPQT 311
 Qy 189 RDVYKA-----NARSGFVSLDVSVTEDLEKSSASHDRVNDYNASLRQESRNKLY 241
 Db 312 APQQQSPHSGSHSRSG-----SGSGSHSSMAS-DGSLRRKSSDLD 353
 Qy 242 RDGKTRIKDTPNGAESHLATENHSQEGHGPS--EID-----NDREYS 283
 Db 354 SHGAQDAQDEEADAFTQSRSERAPEPQLPTKESVDDMLDEVELLGLHSRSGDMSLD 413
 Qy 284 KSRACASLQQLN-EEASDVSDSDMSVDS-----ISSIDVSPD 319
 Db 414 ASPSHSDMMLDKDDVLEDDDDDCVEQKTSQGLCKPKGMDLKKARVENIVSGMRCSPS 473
 Qy 320 DVVGLIGKRFWRARKALANQRFVAVOLF-----ELHRLIKVQKLIASPDLLDEISF 374
 Db 474 SGLAQAGQLQVNGCKKRLKYQFOQHAMERYVAAAAGLNGFLNLSMMLDQEDSESENELES 533

```
QY 375 --LGKVSAPKYPVKLLPS--EFLVKPLPLPHV-----VKQRGDSEKTDQH---KMESSAE 423
Db 534 PQIQQRVEKVALKSQLRSMOEQLAEQKQYVQLCSRMQESECQELDQDVQEQEPD 593
QY 424 NVVGRLSNQHQQSNVMPANPPASPAP-----NGYCFPPQPPPSGNHQOQWLIPVMSPS 479
Db 594 N-----GSSDHIE-----LSPSFTLGDGSDVSNHXKETGQER---PGSSSP 632
QY 480 EGLIYKPHGMAHTGHYGYGYHMPMPVMPVH---PCMGPPPPGNGYFPP---YGM 532
Db 633 SPSPKPKTSLGSSDSGANLWQMSKMSGKLNPLVGVGHAPALPG--FPDLLOHMGD 691
QY 533 MPTIMPY-----CSSQQQQQQPNEQNMQFGHNGFNQNTQQQQQSRD 575
Db 692 MSHAAAMYQQFFFEQEARMAKEAAEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 751
QY 576 NEPAPE--QQQQQPTKSYPRARKSROGSTGSSPSGPO-----GSG----- 613
Db 752 EEOQQQIQRQQQHLQQLQQQQMEQQHVAAPR-PQMHHPAPARLPTRMGGAAGHTALKS 810
QY 614 --SKSPRFAAVIDSDSNINNAPEQITMTT-----TTTTRTIVTQTRDGGGVT 659
Db 811 ELSEKEQMLRA-----NNSSMMWMSGTDLEGLADVLKSEITTSLSALVDII---VT 859
QY 660 RVKVVPHNAKASANAARFISQIERK 688
Db 860 RFV---HORLRFKQADSVTAAAEQLNK 884

RESULT 24
YOKA SCHPO
ID YOKA SCHPO STANDARD; PRT; 964 AA.
AC 074522; Q9UTW9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protein C1494.10 in chromosome III.
GN SPC1494.10 OR SPC170.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Hollroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jegels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neil S., Pearson D., Quayl M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
```

```
RL Nature 415:871-880 (2002).
RN [2]
RP SEQUENCE OF 420-596 FROM N.A., AND SUBCELLULAR LOCATION.
RC STRAIN=968 H90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190 (2000).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC -|- SIMILARITY: Contains 1 Lish domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL023794; CAA19351.1; -.
DR EMBL; AL023776; CAA19308.1; -.
DR EMBL; AB027952; BAA87256.1; -.
DR GeneDB; SPombe; SPC1494.10; -.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
DR GO; GO:0006350; P:transcription; ISS.
DR InterPro; IPR006594; Lish.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS00896; LISH; 1.
KW Nuclear protein.
FT DOMAIN 34 66 LISH
SQ SEQUENCE 964 AA; 104591 MW; B8CC13AD417D66F4 CRC64;
Query Match 4.2%; Score 155.5; DB 1; Length 964;
Best Local Similarity 22.6%; Pred. No. 0.36;
Matches 92; Conservative 59; Mismatches 169; Indels 87; Gaps 20;
QY 234 QESRNLRYDGGKTKLKTGDNGAESHLATENHSQHGCHSP-----EDINDREYSKRACA 289
Db 60 RESKIQPKSSSTAFSPNNAPSPFPKNSL---ASPKISISISGDRLYNHMSAP 116
QY 290 SIQQINE-----EASDDVSDSDMSVDSSISIDVSPDDV-----VGILGQ--KRFWRARKATA 338
Db 117 SPNKKEETNVVHANEIDISLDR-QSFGSSSIPPSSEVINVEGFLVENFNFWDVFSARV 175
QY 339 NQQRVAVOLFE---LHRLIKVQKLIAASPDLLEISFLGKVSASYPVK-KLLPSBFL 394
Db 176 SRVNSTPIQLYDPSRQRMARPMNLQASQVPVPSSTFSRSVAVPNPSLPNPSVLQGVQM 235
QY 395 VKPELPVHVVKQRGDSEKTD---QHKME---SSAENVVGRLSNQHQQHQSNTMPANPP 448
Db 236 NNPTPKGTPTSTIEGAKTS:PPSHAMQNPHNSPPASADRLQKNHPVQSSNFNPYTPAPS 295
QY 449 ASPAPNGYCFPPQPPPSGNHQOQLIPVMSPEGLYKPHGMAHTGHYGYGYHMPMPM 508
Db 296 ITVPN-----YIENTAM-----MGPSVSSFGDTD- 320
QY 509 VMPQVHP-GMGFPPPG--NGYFPYPMPTIMNPYCSSQQQQQQQQQQNQFQHPGNLQ 565
Db 321 --PRTPAGMGPNPTAARNGFYPP---TPAQIH-QLKAQQQHLQFQSKQMSF-PAPINWK 373
QY 566 NTQQQQQ-----RSDNPAPOQQQPKSYPRARKSROGSTGSSPSG 607
Db 374 SNKQQQLQYVDFRGVSGADLQKQWNKS-----TSARGLOPFG 412

RESULT 24
SMF1 HUMAN
ID SMF1 HUMAN STANDARD; PRT; 1902 AA.
AC O14497; Q9UP21;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
```


ID AC SAS DROME STANDARD; PRT: 1693 AA.
 DT 004164; Q960M6; Q9VI73;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative epidermal cell surface receptor precursor (Stranded at second
 DE protein).
 GN SAS OR CG2507
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM LONG).
 RX MEDLINE=92290115; PubMed=1339334;
 RA Schonbaum C.P., Organ E.L., Qu S., Cavener D.R.;
 RT "The Drosophila melanogaster stranded at second (sas) gene encodes a
 RT putative epidermal cell surface receptor required for larval
 RT development";
 RL Dev. Biol. 151:431-445(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).
 RC STRAIN=Berkley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -!- FUNCTION: Vital for larval development.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Name=Long;
 CC IsoId=Q04164-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=Q04164-2; Sequence=VSP_004071;
 CC -!- TISSUE SPECIFICITY: Expressed in most, if not all, ectodermal
 CC tissues which produce a cuticle.
 CC -!- DEVELOPMENTAL STAGE: Throughout development.
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -!- SIMILARITY: Contains 2 VWFC domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

 EMBL; M68866; AAA28879.1; -;
 EMBL; AE003672; AAN13346.1; -;
 EMBL; AE003672; AAF54052.2; -;
 EMBL; AY051979; AAK93403.1; -;
 FlyBase; FBgn0002306; sas.
 GO; GO:0046324; C:apical plasma membrane; NAS.
 GO; GO:0005887; C:integral to plasma membrane; NAS.
 GO; GO:0004872; F:receptor activity; NAS.
 GO; GO:0002168; P:larval development (sensu Insecta); IMP.
 InterPro; IPR008957; FN III-like.
 InterPro; IPR003961; FN_III.
 InterPro; IPR001007; VWFC_C.
 Pfam; PF00041; fn3; 1.
 SMART; SM00060; FN3; 3.
 SMART; SM00214; VWFC; 4.
 PROSITE; PS01208; VWFC_1; 1.
 PROSITE; PS01184; VWFC_2; 1.
 Receptor; Developmental protein; Signal; Transmembrane; Glycoprotein;
 Repeat; Alternative splicing.
 SIGNAL 1 41
 CHAIN 42 1693
 PUTATIVE EPIDERMAL CELL SURFACE RECEPTOR.
 DOMAIN 42 1635
 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 1636 1656
 POTENTIAL.
 DOMAIN 1657 1693
 CYTOPLASMIC (POTENTIAL).
 DOMAIN 51 377
 THR-RICH.
 DOMAIN 534 901
 CYS-RICH.
 DOMAIN 997 1273
 GLU/PRO-RICH.
 DOMAIN 663 708
 VWFC 1.
 DOMAIN 828 902
 VWFC 2.
 DOMAIN 1411 1501
 FIBRONECTIN TYPE-III.
 CARBOHYD 106 106
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 109 109
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 330 330
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 473 473
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 538 538
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 622 622
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 685 685
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 827 827
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 846 846
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	929	929	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	939	939	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1323	1323	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1419	1419	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1517	1517	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	VARSPLIC	930	1274	Missing (in isoform Short).	
FT	CONFLICT	591	591	/FTIG=VSP 004071.	
FT	CONFLICT	1509	1509	V -> L (IN REF. 1).	
FT	CONFLICT	1693	AA;	V -> G (IN REF. 1).	
SEQ	SEQUENCE	1693	AA;	DA50F9667F41DC4	CRC64;
Query Match					
Best Local Similarity 4.2%; Score 153; DB 1; Length 1693;					
Matches 95; Conservative 52; Mismatches 169; Indels 166; Gaps 20;					
QY	157	APSSHSIRPFQENVNQTGSKONVCLATCSK-PEV-RDVQKANARSGGFVISLSDVSVTREI-	213		
DB	919	ATSTHSSIPANETTTTATANKSTISPSVPQIKDEKRPPASGAFTPLDGPKPKSIG	978		
QY	214	-----DLKSASSDRV-----NDYNASLRQESRNLRYDGGTKRLKD-----TDNGA	256		
DB	979	GLGIPEKPEKPEKAHKVKVHQOQQHQOQQEQOQHONDVIFGDRTEEQEELPPNGGF	1038		
QY	257	ESHLATENHSQGEGHGP-----EDIDNDREYSKSRACASLQQINEASDDVDSD	305		
DB	1039	VPFGQGHQPHQLHGYPGYFNVPKPVEDYNPYEY-----DINPGTQP-----	1084		
QY	306	SMWDSISSIDVSPDVVGILGQKRFWRKAIANQRVFAVLFEHLRIKVQKLAAASP	365		
DB	1085	---GKPPPVTSQSDFNLILGAEQ-----	1105		
QY	366	DLLDELISFLGKVSAKSPYVKLLPSEFLVKPLPHVVVKQRGDSEKTQDKMSSAENV	425		
DB	1106	-----PGHPVH-----PGH--CGPPIIHGQTQKNHNHNGFQVRIEQLIQ--	1143		
QY	426	VGRLSNQGHHSOOSNYMFANNPPASPA PNGCYCPPOPPSGSNHQOQLIPVMSPEGLLYK	485		
DB	1144	-----HLQQT-----VPGG-----PPPPP HQHQHS-LTPOLHPQQQIQS	1178		
QY	486	PHPGMAHTGHYGYYGHMYMPVMVMPQVHPGMGFPFPGNGY-----FPYGMVPTI	536		
DB	1179	QHP-----GHVPI-----VHSGVPPPPPHGHGIAIVDGTVAYESYVPVPGI	1220		
QY	537	MNPYCSSLQQQQQQPNQNMQNFQHG-----NLQNTQQQQRSNDNEPAPQQQQPTKS	589		
DB	1221	GVPQ-HHFQQHQTTPQCHLQQTILPSSSTTSLGSTQASEHSHQNOKLARQQSQGANLI	1279		
QY	590	YP	591		
DB	1280	QP	1281		

Search completed: July 29, 2004, 08:22:44
Job time : 23 secs

Thu Jul 29 08:52:25 2004

```

90 161.5 4.4 530 10 Q9LYK5
91 161.5 4.4 1169 11 Q8CHB3
92 161.5 4.4 1263 11 Q8OYB2
93 161.5 4.4 3498 5 Q20497
94 161 4.4 1083 16 Q86637
95 161 4.4 1146 3 Q8XOL3
96 161 4.4 1322 5 Q9VVH2
97 160.5 4.4 508 10 Q9M0H8
98 160.5 4.4 1441 4 Q00150
99 160 4.4 699 10 Q8LQU0
100 160 4.4 752 11 Q8R506

```

ALIGNMENTS

RESULT 1

```

Q9SNQ6 PRELIMINARY; PRT; 760 AA.
ID Q9SNQ6
AC Q9SNQ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 24, Last annotation update)
DE Similar to Arabidopsis thaliana chromosome II BAC T19L18.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
clone P0535G04."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP000399; BAA83571.1; -.
DR Gramene; Q9SNQ6; -.
DR GO; GO:0006364; P:RNA processing; IEA.
DR InterPro; IPR000238; Rib_bind_facta.
DR PROSITE; PS01319; RBFA; I.
DR SEQUENCE 760 AA; 82378 MW; 253BP51P93D35670 CRC64;
SQ

```

```

Query Match 21.1%; Score 773.5; DB 10; Length 760;
Best Local Similarity 32.1%; Pred. No. 3.4e-46;
Matches 261; Conservative 105; Mismatches 257; Indels 191; Gaps 36;

```

```

QY 4 GKDEE-KILEMPFRLHVNDAK-GGPRAPPRNKMALEYQLSIPSRF---GDHGTWNS 57
DB 12 GKEARKVMGLFRLHVNDAKGGGPRAPPRNKMALEYQFTVPSHRFSGGGGGVGG 71
QY 58 RSNNTSTLVHPGSPQCGVERNL-----SVQHLDSAAANQATEKFVQMSF 104
DB 72 PAHSTSAASQSQSQSVYGRDSSLFQPPNVFSPNRPGHSTKINSKINKISGRKELGM 131
QY 105 MEN-----VRSSAQHQR-----KVMREEDPAPVYINSRSSQHGRTK 144
DB 132 LSSQTKGMDIYASRTAEAPQKRAENTIKSSSGKRLADDDDFMVPFVSFNRFPQSTQEN 191
QY 145 SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV-CLATCSKPEVRDQVK-ANARSGGFV 202
DB 192 AGVQ-DQSTPLVAAMPKHS-----PSTVSKSTKCYNTVSKLRIHVSVDKSTPLKD 244
QY 203 ISLDVSVT-EETDLEKSSASHDRVNDYNASLRQESRN-RLYRDGGKTR-LKQTD---NGA 256
DB 245 KEMEAQTSKNVEVEKSSPHASKDMF-----ESRHAKVYPMKDKTGIIINDSEPHGN 298
QY 257 ESHLATENHS-----QEGHSGPEDTDNDREY-----SKSRACASIQLIN 295
DB 299 SGHQATSRNGSGMKQNPMPMRNRNEISSNPSSSENTDRHYNLPQGGIEETGTKRKLLEQHD 358
QY 296 EFASDDVS-----DDSNVDSISIDVSPDDVVGILGQKRFWRKARKAIANQO 341

```

```

Db 359 AEKSDVDSKLLQHQDAENIDVDSSVREICITGWEISPDKIVGAIQTGKHFWKARRAIMNQO 418
QY 342 RVFAVOLPELHRLIKVOKLIAASPOLLDDEISFLGKVSASYPVKKLIPSEFLVKPPLPH 401
Db 419 RVFAVQVPELHRLIKVOKLIAASPHVLIESDPCIGNALLGS---KNKLVEENLKAQPLLV 475
QY 402 VVVQKRGJSEKTDQHKMSESAENV-----VGRLSNQGHQHQSNVMPFAN---NPSA 449
Db 476 ATI-----DDVESLQPEVSKENTEDSPSPHDTGLGSGQRDQAATNGVSKNRRATPYA 531
QY 450 SP-APNGYCFPPPPSPGHNQOWLIPVMSPSGLIYKPHPG-----M 490
Db 532 SDNKQNNWGVQLQPP-----QNQWLVEVMSPLEGLVYKPYSGPCPPAGSILAFYANCTPL 587
QY 491 AHTGHYGVYGHYMPPTVMYQVHPGMGPPP--PGNGYPPPYGMMPTIMNPYSSQSQOQO 548
Db 588 SLFSTAGDFMNSAYGVMPHPQHQHMGAPGPPSPMNM-YFPFPSI--PVNPP----- 635
QY 549 QQFNEQMNQGHFGNLQ---NTQQQQQRSDNBPAPQOQQOQPTKSY-----RARKSR--- 597
Db 636 TAPAPVVEQGRHPSMPQPYGNFEQQSWISCN-----MSHPSGIWRPHASRDEA 684
QY 598 QGSGTSSPSGPGQISGSKSFRPAPAAVDEDSNINNAPETMTTITTTTTRTIVTQTTDGGG 657
Db 685 QASSASSPDRFQCSGS---GPVSAPPTVSAQNNQPOP-----SYSSRD--N 726
QY 658 VTRVIVKVPENAKLASENAAIRIFQSIQERKRYD 691
Db 727 QTNVIVKVPHNSTRASAAIRIFRSIQMERQRDD 760

```

RESULT 2

```

Q942Q5 PRELIMINARY; PRT; 765 AA.
ID Q942Q5
AC Q942Q5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P697C12.15 protein.
GN P697C12.15.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone P0697C12."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003296; BAB64680.1; -.
DR Gramene; Q942Q5; -.
DR SEQUENCE 765 AA; 82653 MW; 5CE1C45576D4D83C CRC64;
SQ

```

```

Query Match 19.2%; Score 704.5; DB 10; Length 765;
Best Local Similarity 29.5%; Pred. No. 2.6e-41;
Matches 244; Conservative 113; Mismatches 263; Indels 207; Gaps 38;

```

```

QY 4 GKDEE---KILEMPFRLHVNDAK-GGPRAPPRNKMALEYQLSIPSRF-GDHGTWNSR 58
DB 8 GKEVEERKVMGLFRLHVNDAKGGGPRAPPRNKMALEYQFTVPSHRFSGGGGALASA 67
QY 59 SNNVSTLVHPGSPQCGVERNL-----SVQHLDSAAAN----- 92
DB 68 RGLSTARSTSAASQSQSVYCGDMPLEFPFNVPSNGPQSQVEKNNSNVNPNQINGSRKDSGML 127
QY 93 ----QATEKFVS----QMSFMENVRSSAQHQKRMVREEDPAPVYINSRSSQHGRTK 144
DB 128 STQPKGIDYSGSRAECAPQQRVEKIGKSSGRKLADDDDFIVPSFARSFPYSTKER 187
QY 145 SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSK-----PEVRDQVKANAR 197

```


RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,
RA "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.",
RT Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115682; AOS2645.1; -;
DR InterPro; IPR003100; PAZ;
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; Piwi; 1.
SQ SEQUENCE 1208 AA; 137164 MW; B208B969B71CE89C CRC64;

Query Match 5.5%; Score 200.5; DB 5; Length 1208;
Best Local Similarity 24.5%; Pred. No. 2.5e-05;
Matches 81; Conservative 28; Mismatches 100; Indels 121; Gaps 13;

QY 386 KLLPSEFLVKPLPHVVKVQKQDSEKTDHKKMSSAENVVGRSLNQGHHQQSNYMPFAN 445
Db 8 RDLPLSD-----DDGDLNR-KSRDENIRHPQOQOQOQOQOQOQOQOQOQ 54

QY 446 -----NPPASAPNGYCFPPPPPPSGNHHQWLIPVMSPEGLYKPHFGMAHTGHYGG 498
Db 55 YQOQPVDSQSSQGGYQVPPPPPSGYQ-----SP-----PPYSRAGY--D 96

QY 499 YGHYMPTMVMQYHFGMGFPFPGNGYFPYPPYGMPTMMPYSSQOQOQOQOQNEQMNQF 558
Db 97 PYGYAPP-----FQHQNSYVPOQS--FDYFQQYQPPPPPPQOQOQOQYQPSOQQOY 150

QY 559 -----GHCNLQNTQQOQSDNEPAQO-QOQPTKSYPRARKSR 597
Db 151 QPOQSHGYDYQSQYQYQNOQOQOQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 210

QY 598 Q-----GSTGSSPSGPQ-----GIGGS----- 614
Db 211 QOQDIFQVQDNRYSYRSDGDRGSGSGNDNYGYQGRGRDNNNNNNGGSGSGDRYG 270

QY 615 -----KSRFPFAAVDSDSNIN 631
Db 271 SGSGSGGDRYSGRGRDRDFGDRDNNNNNN 300

RESULT 5
QY Q9V6P4 PRELIMINARY; PRT; 1503 AA.
AC Q9V6P4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG4744 protein.
GN CG4744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Ii P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,


```

Db 411 DHGTNSSSTFTSSAKRPFQTSRGD--MYNDNNGAGYKKSRRHTVSCNLFVKKRTADTYAI 468
QY 263 E--NHSQEGHSPEDIDNDREYKSRACASLQOINEEASDDVSDSDVSDSISSIDVSPD 320
Db 469 EVFNRFRDGTG---LETDIMFLKPRMELG-KLINDAAYNGVGVVNVNKNTHNDV--- 519
QY 321 VVGILGQKRFWRARAKAIAQQRVAVQVLFELHRLIKVKQLIAASPDLLILDEISFLGKUSA 380
Db 520 ---QTFYKGSQGET-----KFDEYIISAD---DAVAIFNNIK- 551
QY 381 KSYPVKLLPSEFLKPPPLPHVVVVKRGDSEKTDQHKMESSAENVVGRLSNQHGHQSNY 440
Db 552 -----NNRNSRPTDTRAMS-----HOQNIY 572
QY 441 ----MPFANNPPASAP-----NGYCFPP-----QPPPSGNHQWLIPVMSPEGLIYKPH 487
Db 573 GAPPPLVPNGAVGPPPTQNYGYSMPPPPQOQQOQPYNGY---MPPPSHDQG---YGSQ 627
QY 488 PGMAHTGHYGYGYHYMPTMVPMPQYHGMGF-----PPPGNGYFPP----- 529
Db 628 PPPIPMQSYGRYGTSTIPPPP---PQQQIPQGYRYQAGPPPO---PPSQTPMDQOQLLS 680
QY 530 ---YGMPTTMMNYCSCSQOQQOQQOQNEOMNOFGHPGNLQNTQQOQORSNDNEPAPQOQQOQPT 587
Db 681 ATQNLPPNVNLLSMAOQQOQQOQPHAQOQLV---GLIQSQGQ-----APQOQQOQL 729
QY 588 KSYPRARKSQGSTGSSPSGPGISGSKSRPFA 621
Db 730 GGYSSMNSPPPMSTNYNG-QNISAKPSAPPMWS 762

RESULT 8
Q81HHO PRELIMINARY; PRT; 1500 AA.
AC Q81HHO;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE AT02321P.
GN CG4744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BT001259; AAN71015.1; -.
DR FlyBase; FBgn0033834; CG4744.
SQ SEQUENCE 1500 AA; 170871 MW; 1AD9D6686FEC90C0 CRC64;

Query Match
Best Local Similarity 22.6%; Score 189.5; DB 5; Length 1500;
Matches 148; Conservative 74; Mismatches 247; Indels 187; Gaps 33;

QY 4 GKDEKILEPMPF-----RLHVNDADKGGPPAPPFNKALYEQIIPSQRGDHGTMSR 58
Db 398 GMDGNRMME--LPQSYDYRVHGYD-----RFHPLNQ-----RQMS-----NVGDY-----ER 437
QY 59 SNNVSTLVHPGSSPCGVERNLSYQHLDSAAQAQATEKFVSQMSFMENVRSSAHDQK 118
Db 438 SNR-----MPLDLRFKRPHESEFLEAPNELSLGQLSVPEPVLKSEDEN--- 483
QY 119 MYREBEDFAPVYINRSRQSGHRTKSGIEKHKHTPMVAPSSHHSIRFOEVNQTSQKNV 178
Db 484 -ISDHAMKIAVLIDIKDKALHSTGDGESKEKT-----LEEVKLEGLHEEL 530

```

```

QY 179 CL--ATCSKPEVRDOVKANARSGGFVILDSVSVTEIDLEKSASSHDRVNDVNASLRQES 236
Db 531 KLDESLVKPSKLDSEVSCINGKF-----EDRSEMEVENKPLGPVK-REAKAELINQM 583
QY 237 RNRLYRDGGKTELKDTDNGAESHLATENHSQBCHGSPEDIDNDREYSK-SRACASIQOIN 295
Db 584 GNKLKNDGSTLVAK---SSAESPLKEDVPPNLAARESELNVSTGTDLASVSKKONSLKRRK 640
QY 296 EFASDDVSD--DSMYD-----STSSIDVSPDDVVGILGQKRFWRARAKAIAQQRV 343
Db 641 SSATSDSDSDKHSNPDCTECKSEDAEKSETDAEKD----- 677
QY 344 FAVQFELHRLIKVKQLIAASPDLLILDEIS-FLGKVSASYP-----VKLLPSFLVK 396
Db 678 ---SIVKNIKLQLOEPVEEKSE---CEEISKVVGKGAASAPKVNPDAAVVKLIANLSMD 732
QY 397 PPLPHVVVKQSG-----DSEKTDQHKMESSAENVVGRLS---NQGH 435
Db 733 -----KDRGKQARHRRYRRQRNRKRSLSDNDQHSDDKQHAVSKRLAPLLDDFED 784
QY 436 QOSNYMPPFANNPPASAPNGYCFPPQPPSPSGNHQWLIPVMSPEGLIYKPHGMAH--- 492
Db 785 HNGNRVGFQN-----LGNG-MLTFDAEYSGERS-----YPSHEMLAHFDE 823
QY 493 ---TGHYG-GYGYHYMPTMVPMPQYH--PGMGPPPPGNGYFPPPYGMPTTMMNYCSCSQOQ 547
Db 824 LETAHGDSMWPIYDPTKVVVPMVNRPNNGF-----GQOQPVQM---FSQOQP 870
QY 548 QOQNFQMNQ-----FGHPGNLQNTQQOQORSNDNEPAPQOQQOQPTKSPRARKSQGS 600
Db 871 QOQMQOQOQHMQOQPPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 925

RESULT 9
Q86HN1 PRELIMINARY; PRT; 1165 AA.
AC Q86HN1;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to PH (Plectstrin homology) domain.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellum.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=AX4;
RC MEDLINE=22092622; PubMed=12097910;
RX Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,
RA "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=AX4;
RC Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116957; AA052542.1; -.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00100; DH_2; 1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS01178; ZF_FYVE; 1.

```


OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
[1]	SEQUENCE FROM N.A.
RN	Hudson J.B., Podos S.D., Simpson S.L., Keith K., Ferguson E.L.;
RP	Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RL	EMBL: AF039232; AAC09260.1; -.
DR	HSSP: Q13485; 1DD1.
DR	FlyBase: FBgn0011655; Med.
DR	GO: GO:0005737; C:cyttoplasm; IDA.
DR	GO: GO:0005634; C:nucleus; IGI.
DR	GO: GO:0003702; F:RNA polymerase II transcription factor acti. . ; NAS.
DR	GO: GO:0008283; P:cell proliferation; IMP.
DR	GO: GO:0009950; P:adult ventral axis specification; IMP.
DR	GO: GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR	GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR	GO: GO:0007179; P:IGFBeta receptor signaling pathway; IGI.
DR	GO: GO:0007476; P:wing morphogenesis; IMP.
DR	InterPro: IPR001132; Dwarfin.
DR	InterPro: IPR003619; Dwarfin A.
DR	InterPro: IPR008984; SMAD_FH_A.
DR	Pfam: PF03165; MH1; 1.
DR	Pfam: PF03166; MH2; 1.
DR	SMART: SM00523; DWA; 1.
DR	SMART: SM00524; DWB; 1.
DR	SCOPED: SCOPED_745_AA; 79071 MW: 316FOD71F32B7E9D CRC64;

DR	SMART; SM00324; DWB; 1.	SEQUENCE	745 AA; 79071 MW; 316F0D71F32B7E9D CRC64;	
	Query Match	5.1%;	Score 186; DB 5; Length 745;	
	Best Local Similarity	22.8%;	Pred. No. 0.00014;	
	Matches	109; Conservative	56; Mismatches 174; Indels 140; Gaps	22;
QY	265	HSQEGHSGPEDIDNDREYSKRCASIQINERASDDVSDSMVDSSTSIDVSPDDVGI	324	
DB	28	HRQG-----ESEGFAKRAJTESLVKKLKERDEL--DSLTAITTINGAHPKSCVTVI	76	
QY	325	LGQKRFWRARKAIANQ---RVFAVQLF--ELHR-LIKVKOKLAASPDLLLDLSEIF---	374	
DB	77	--QRTIDGRLLQVAGRGFPVIVARWMDPLHKLHKLHVKYCAFAFDLKDCSVCVNPY	133	
QY	375	-----LGKVSAKSYVVKLLPSEFLVKPPLPHVHVVKRGDSEKTDQHKMESA	423	
DB	134	HYERVVSGIDLSGLSLSQGP-SRLVKDEYSAGPLVGS-----DIDNDIGITQHHT	186	
QY	424	NVGRSLNQGHQ--QSNYMPFANNPPASPANGYCFP---POPSPGHQWLLP-----	474	
DB	187	QMVGFP-GGYGYPQGPSEYVGDA-NPMSAMFTGRTPIKIEPDGVAGSRGSMVPPPPPL	244	
QY	475	-----VMSSEGLLYK---PH--PGMAHTGHVGYGYGHYMPM	508	
DB	245	GQPPQQQQQQPQQQTPOPTQQQAQSAHAHSLPVPHGMPGMP-----GPMNPGV	295	
QY	509	VMPQVHPGMGPPPCNGYFPYPYGMPTMINPYCSSQQQQQQQQPNEQMNQFHGPNLQNTQ	568	
DB	296	WAPPPPPQQAQNPQNGVHTHTQANSPTDPASALAMQQQQQQQQQQQQQQ-----Q	345	
QY	569	QQQQRSD-----NEPAPQQQQQ-----PTKSYPRARK-----SR	597	
DB	346	QQQQQSGGVPGNSVNAGGGAAGGYGQPPVPSQMGAGGGGTSVAPSVHAQQNGYVSQ	405	
QY	598	QGSTGSPSPGQIGISGSKSRFFAAVDEDSNNINNAPEQMTTTTTTTRTVTTQTTDGG	656	
DB	406	PGSAGSAPVGGGVGTGAQFTP-----QQPQDPTGVQANTSGAQAAGAGGG	453	

RESULT 13	
QBIMG5	
ID QBIMG5	PRELIMINARY;
AC QBIMG5;	PRT; 745 AA.
DT 01-MAR-2003	(T=EMBLrel. 23, Created)
DT 01-MAR-2003	(T=EMBLrel. 23, Last sequence update)

Thu Jul 29 08:52:25 2004

RA 01-AUG-1998 (TrEMBLrel. 07, Created)
 RT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MED protein (ID22279p).
 GN MED OR MEDEA OR CG1775.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Mecoptera; Arthropoda; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.D.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brattier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davies P., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
 RA Jalali M., Kalish J., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson J.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris L., Muzny D.M., Nelson D.L.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Nuskern D.R., Paclab J.M.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Scheeler F., Smith T.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Strong R., Wang X.,
 RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Swirskas R., Tector A.C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CN BW;
 RX MEDLINE=98171467; PubMed=9502733;
 RA Das P., Maduzia L., Wang H., Finelli A., Cho S.-H., Smith M.,
 RA Padgett R.;
 RT "The Drosophila gene Medea demonstrates the requirement for different
 RT classes of Snads in dpp signaling."
 RL Development 125:1519-1528 (1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CN BW;
 RX MEDLINE=98171458; PubMed=9502724;
 RA Wisotzkey R.G., Mehra A., Sutherland D.J., Dobens L.L., Liu X.,
 RA Dohmann C., Attisano L., Raftery L.A.;
 RT "Medea is a Drosophila Smad homolog that is differentially required
 RT to potentiate DPP responses."
 RL Development 125:1433-1445 (1998).
 RN [4]
 RP SEQUENCE OF 27-771 FROM N.A.
 RX MEDLINE=98359837; PubMed=9693372;
 RA Inoue H., Imamura T., Ishidou Y., Takase M., Udagawa Y., Oka Y.,
 RA

RA Tsuneizumi K., Tabata T., Miyazono K., Kawabata M.;
 RT "Interplay of signal mediators of decapentaplegic (Dpp): molecular
 RT characterization of mothers against dpp, Medea, and daughters against
 RT dpp.";
 RL Mol. Biol. Cell 9:2145-2156 (1998).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF041433; AAC38972.1; -;
 DR EMBL; AF027729; AAC38971.1; -;
 DR EMBL; AF057162; AAC62005.1; -;
 DR EMBL; AY069496; AAL39641.1; -;
 DR HSSP; Q13485; 1DD1.
 DR TRANSFAC; T04379; -;
 DR FlyBase; FBgn0011655; Med.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0005634; C:nucleus; IGI.
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti...; NAS.
 DR GO; GO:0003283; P:cell proliferation; IMP.
 DR GO; GO:0009280; P:cell division; IMP.
 DR GO; GO:0007480; P:regulation of transcription, DNA-dependent; NAS.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 DR GO; GO:0007179; P:regulation of transcription, DNA-dependent; NAS.
 DR GO; GO:0007476; P:wing morphogenesis; IMP.
 DR InterPro; IPR001132; Dwarfin.
 DR InterPro; IPR003619; Dwarfin A.
 DR InterPro; IPR008984; SMAD_FHA.
 DR Pfam; PF03165; MHL1; 1.
 DR Pfam; PF03166; MH2; 1.
 DR SMART; SM00523; DWA; 1.
 DR SMART; SM00524; DWB; 1.
 SQ SEQUENCE 771 AA; 81638 MW; 8E60AB1E4F4946B3 CRC64;
 Query Match 5.1%; Score 186; DB 5; Length 771;
 Best Local Similarity 22.8%; Pred. No. 0.00014;
 Matches 109; Conservative 56; Mismatches 174; Indels 140; Gaps 22;
 QY 265 HSEGHGSPEDIDNDREYSKRCASLQIQNEASDDVDSDMSVDSISDVSDDVGI 324
 Db 54 HRQGG-----ESGFAKRAESLVKLEKRDDEL--DSLITAITTNGAHPSCVTI 102
 QY 325 LGQKRFWRARAIANQQ---RVFAVQLF---ELHR-LIKVQKLTAAASPDLLDLSIF--- 374
 Db 103 ---QRTLDGRQLVAGRGFPFVIYARIWRWPDLLHKLKHVKYCAFAFDLKCDSCVNPY 159
 QY 375 -----LGVSAKSVKVLPLSEFLVLPVHVVKRGDSEKTDQHKMSSAE 423
 Db 160 HYERVSPGIDLSGLSLQSGP-SRLVKDEYSAGPLVGSN-----DIDNDIGITQHTPT 212
 QY 424 NVVGRLSNQHQQ-QSNMYPFANNPPASPAPNGYCFP---PQPPSGNHQOWLP----- 474
 Db 213 QMVGP-GGYGYPQGPSEYVGD-A-NFMSAMFTGRTIPKIEPQDGVAGSRGSMVPPPR 270
 QY 475 -----VMSFSEGLIYK---PH---PGMAHTGHYGYGYHYMPTPM 508
 Db 271 QGPPQOQ 321
 QY 509 VMPQYHGMGFPFPPGNGYFPYGMPTIMNPFYCSQQQQQQQQQQQQQQQQQQQQQQQQQQ 568
 Db 322 MAPPPPOAQNPQNGVHHTQANSPTDPASALAMQOQOQOQOQOQOQOQOQOQOQOQOQOQ 371
 QY 569 QQQQRSD-----NEPAPQOQ 597
 Db 372 QQQQSGGVGPNVAGGAAGGQYQPPVPSQMGAGGGGTSAVPSVHAQQNGYVSQ 431
 QY 598 QGSTGSSGPGQIGSGSKSFRFAADEVSDSNINNAPEQTMTTTTTTTTTTTTTRTQTTRDGG 656

```
Db 432 PGSAGSAPVGGGVGTGTAQTP-----OOQOQPPPTGVQANTGSAGAGAGGG 479
RESULT 15
Q9QY60 PRELIMINARY; PRT; 756 AA.
ID Q9QY60
AC Q9QY60
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cgl/Xap80 protein.
GN Cgl/XAP80.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platzer M., Zhao W., Herman G.E., Rosenthal A.;
RT "Comparative sequence analysis of the mouse Mtm locus and the
RL corresponding region of human Xq28.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF125313; AAF22120.1; -.
SQ SEQUENCE 756 AA; 81963 MW; 8F5B672B378A5943 CRC64;

Query Match 5.0%; Score 184.5; DB 11; Length 756;
Best Local Similarity 21.6%; Pred. No. 0.00018;
Matches 106; Conservative 70; Mismatches 179; Indels 135; Gaps 22;

QY 261 ATENHSQEGHSPEDINDREYKSRACASLOQINEASDDVSDSDSISIDVSPDD 320
Db 81 STADTGLKGAVP-----YFEKN--SMPADQELQDLLEELTKIQEPSSNDL---D 127
QY 321 VVGILGQKFRWARKAIANQORVFA-----VOLFEHLRLIKVKQLIAASPDLLDEISF 374
Db 128 LEKILGSKP--EFPLVLHNPQAPLGPAPKLVPQMPHESLGSKSEFASCSQVAGTSLPI 185
QY 375 LKVKSAKYPV-----KKLPS-----EFLVKPPLPHVVKQKGDSEKTDQHK 417
Db 186 MPSTGMSYIPSSSKQIVSSSSSTAQAQVKQVQNNLPTVMPPLSVPQWHA-----HQ 240
QY 418 MESSAENVVGRLSNQHQQ--SNMPPANNPPASPANGVCFPPQPPP-----464
Db 241 LKALAASKGATKQGSNRNWSLPPGLSPPLVPSPHPPPPPPPPPCSPQNFATSC 300
QY 465 -----SGNHQ-----WLIPVMSPSRG-----LIYKPH--PGM-----490
Db 301 MSSSLSGSAQSSPNALLSSMAPSSNASLPTLPYVPAPKLPGPLNQOPQPSQSSILA 360
QY 491 -AHTGHYGGYGHY-----PTPMVMPQYHPGMGFP--PGNGYFPPYGMMPPTI--MNPY 540
Db 361 NLVSSSVKSPQGHLSALPTSTPGSPFYPENLSPLGLPQGSFTFYSLRSLRSLTPTSL 420
QY 541 CSSQQQQQQQNEQNMNFGHPGNLQNTQQQQQSDN---EPAPQQQQQTKSPRARKR 597
Db 421 LSQQQQQQQQQQQQQQQ-----QQQQQQQQQQHANSIFKPMTSSQQPKTSL-----MIMQ 471
QY 598 QGSTGSSPSGPQ--GLSGSKSFAPPAVDEDSNI-----NNAPE 634
Db 472 QGLSSSSPEAPEPFTNTKPLSHFVSEPSQKMSMSTHSRQSSLLHYLPQATPAHAPS 531
QY 635 QNTWTTTTTTT 644
Db 532 ATASSTATAT 541

RESULT 16
Q819J8 PRELIMINARY; PRT; 960 AA.
ID Q819J8
AC Q819J8
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
SR-related CTD associated factor 6.
SCAF6 OR CG6615 OR CG6626 OR CG14058 OR CG32168.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
STRAIN=Oregon R;
Sampson N.D., Hewitt J.E.;
"Functional characterization of the novel SR-related CTD associated
factor, SCAF6.";
Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
EMBL; AF36543; AAN77184.1; -.
Flybase; FBgn0052168; scaf6.
GO; GO:0005622; C:intracellular; IEA.
GO; GO:0003723; F:RNA binding; IEA.
GO; GO:0006396; P:RNA processing; IEA.
InterPro; IPR000467; G_patch.
InterPro; IPR000061; Surp.
Pfam; PF01585; G_patch; 1.
Pfam; PF01805; Surp; 1.
SMART; SM00443; G_patch; 1.
SMART; SM00648; SWAP; 1.
PROSITE; PS01174; G_PATCH; 1.
PROSITE; PS01128; SURP; 1.
SEQUENCE 960 AA; 107559 MW; CE968612D75EB72F CRC64;

Query Match 5.0%; Score 183.5; DB 5; Length 960;
Best Local Similarity 20.3%; Pred. No. 0.00029;
Matches 128; Conservative 67; Mismatches 203; Indels 231; Gaps 27;

QY 78 ERNLSVQHLDSSAANAQATEKFVSQMSFMENVRSSAQHQRKMYREEDFAVPVINSRRS 137
Db 172 EANLSAQH--TALMNQTKQIEDAIAAQAQEQMAGEQIVLDFDGVLIPIESCCK 229
QY 138 QSHGRTKSGI-----EKEKHTPMVADPSHHSTRPQE 168
Db 230 DSIASGKNILQHSSTDSAKINVLVLLKALVNGSTFOQKLHLIYLVNDILHCHMKRNI 289
QY 169 VNQTQKQNVCLAT-CS-----KPEVRDOVKANARSGGFVISLDVSVTEID--LEKSA 219
Db 290 NDLKNSLENNVPMFCSDLIQNPQOKLAK-----LLSLWESKAKFFDCVSKLQ 342
QY 220 SHDRVNDYNASLROBSRNRLYDGGKTLKDTNGAESHLATENHSQEGHSPEDIND 279
Db 343 SPDSSMOEY-----KTNLQNT-----HHDITAKFTQNTKALDNYQK 380
QY 280 REYSKSRACASLQIQINEASDDVSDSDSISIDVSPDDVVGILGQKFRWARKAIAN 339
Db 381 HHVFTQHASQQIVQLEKQAOH-----L 402
QY 340 QQRVFAVOLFEHLRLIKVKQLIAASPDLLDEISFLGKVSASYPVKVLLPSEFLVKPPL 399
Db 403 EQQIVAKSQQQQHMLOHQQKNI---PSLSMRHIA-----M 434
QY 400 PHVVVKQGDSEKTDQHKMWSAENVVGRLSNQHQQHQSNNYMPFANNPPASPAGYCFP 459
Db 435 P-----GDHE-----HMNSQGHDDQH--PQAGN---NMYPGN-PQ 465
QY 460 POPPPSGNH-----QQWLI PVMSPSGLIKYKPHPGMAHTGHYGGYGHYMTMVMPOY 513
Db 466 QQQPPPGNHFDPRGFNFIPDMS-----KPPFG-----FPGPMHNNQFAMGQQQP 512
QY 514 HPGMGFP--PGNGYFPPYGMMPITMNPYCSSQQQQQQQNEQNMNFGHPGNLQNTQQ 570
Db 513 QPQQ--FPWQAPGSGNEPVDLGLVLSAALQACLQWQQHHPDDPQQQGGH--QLQQQQQQ 569
QY 571 QQRSDNEPAPQQQQ--PTKSYPR--ARKSRQGSTGSSPSGFGQISGSKSRFPFAVDE 626
Db 570 Q-----PQQQKQPPMAAYPQNI VPRPGQNADSS-----EVD 601
```


RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J.A., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Shou S.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Shapleton M., Strong R., Svirska R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kanink J.S., Prochink S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Beniker S.E., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A5003779; AAN14279.1; Med.
 DR FlyBase: FBgn0011655; Med.
 DR GO: GO:0005737; C:cytoplasm; IDA.
 DR GO: GO:0005634; C:nucleus; IGI.
 DR GO: GO:0003702; F:RNA polymerase II transcription factor acti. . .; NAS.
 DR GO: GO:0008283; P:cell proliferation; IMP.
 DR GO: GO:0009950; P:dorsal/ventral axis specification; IMP.
 DR GO: GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 DR GO: GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
 DR GO: GO:0007476; P:wing morphogenesis; IMP.
 DR InterPro: IPR001132; Dwarfin.
 DR InterPro: IPR003619; Dwarfin A.
 DR InterPro: IPR008984; SMAD_FHA.
 DR Pfam: PF03165; MH1; 1.
 DR Pfam: PF03166; MH2; 1.
 DR SMART: SM00523; DWA; 1.
 DR SMART: SM00524; DWB; 1.
 SQ SEQUENCE 671 AA; 71127 MW; 3B5F373690F434FC CRC64;

Query Match 5.0%; Score 182; DB 5; Length 671;
 Best Local Similarity 24.3%; Pred. No. 0.00022;
 Matches 107; Conservative 42; Mismatches 155; Indels 136; Gaps 21;
 QY 265 HSQEGHSGPEDIDNREYKSKRACASLQINFEASDDSDMSVDSIDVSPDDVGI 324
 DB 28 HRQGG-----ESEGFAKRAIESLVKLEKREDEL--DSLITAITNGAHPKCVTI 76
 QY 325 LQKQKFRWARKAIANQ---RVFAVOLF---ELHR-LIKVQKLIASPDLLDEISFLGK 377
 DB 77 ---QRTLDGRQLVAGKGFPHVIYARIWNPDLHKNELKHVYCAFAFDLKCDVC--- 129

QY 378 VSAKSYPVKLLPSEFLVKPLP-HVVVKQRGDSEKTDQHKMSSAENVVGRLSNQG--- 433
 DB 130 VNPYHY-----ERVVSGIDLSGLSLOSGPRLV---KDEYSAGPLVGSMDIDGNDI 178
 QY 434 ---HHQSQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQWLIPVMSPSGLYKPH-- 487
 DB 179 GTIQHH-----PTQWVGPGGYGYP-----OGPSYAV--PHGM 209
 QY 488 PGMAHTGHYGYGYHYMPTPMVPMQYHPGMGFPFGNGYFPFYGMMPETIMPYCSSQQQQ 547
 DB 210 PGMP-----GPMNPGFVMAAPPPPPQQAQNPQGNVHHTQANSPTDPASALAMQQQQ 260
 QY 548 QQQPNEQMNQFGHGNLQNTQQQQQRSD-----NEPAPQQQQQ-- 585
 DB 261 QQQQQQQQQQ-----QQQQQQGGVNGSVNAGGAAAGGYQYGPVPSQMCGA 310
 QY 586 ---PTKSYPRARK-----SRQGSTGSSPSGPGQISGSKSRFRPFAAVDEDSNINNAPEOT 636
 DB 311 GGGGTSVAPSVHAQONGYVQPSGASAPVGGGVFGTAQPTP-----QQPQOP 359
 QY 637 MTTTTTTRTIVTQTRDGG 656
 DB 360 PTGVQANTGSAGAQAAGAGGG 379

RESULT 19

Q8IMG6

ID Q8IMG6 PRELIMINARY; PR: 697 AA.

AC Q8IMG6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE CS1775-FB.
 GN MED OR CG1775.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;

RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

Thu Jul 29 08:52:25 2004

SVIRSKAS R., Tector C., Turner R., Venter E., Wan K.H., Holt R.A., Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodgett, Worley K.C., Yao S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhang M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*."
 Science 287:2185-2195 (2000).
 [2]
 SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y., Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A., Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M., Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D., Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A., Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J., Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A., McIntosh T.C., Moy M., Murphy B., Patel S., Pfeiffer B., Pacleub J., Paragas V., Park S., Puri V., Richards S., Scheeler F., Phouanavong S., Pittman G.S., Rubin G.M., Tector C., Tyler D., Stapleton M., Strong R., Svirska R., Tector C., Tyler D., Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 "Sequencing of *Drosophila melanogaster* genome."
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D., Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E., Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N., Kronmiller B., Marshall B., Millburn G., Richter J., Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 "Annotation of *Drosophila melanogaster* genome."
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 SEQUENCE FROM N.A.
 RA FlyBase;
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003779; AANI4277.1; -;
 DR FlyBase; FBgn0011655; Med.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0005634; C:nucleus; IGI.
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; NAS.
 DR GO; GO:0008283; P:cell proliferation; IMP.
 DR GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
 DR GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 DR GO; GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
 DR GO; GO:0007476; P:wing morphogenesis; IMP.
 DR InterPro; IPR001132; Dwarfin.
 DR InterPro; IPR003619; Dwarfin A.
 DR Pfam; PF03165; MH1; 1.
 DR Pfam; PF03166; MH2; 1.
 DR SMART; SM00523; DWA; 1.
 DR SMART; SM00524; DWA; 1.
 SQ SEQUENCE 697 AA; 73679 MW; 222398AA59F1D28E CRC64;
 Query Match 5.0%; Score 182; DB 5; Length 697;
 Best Local Similarity 24.3%; Pred. No. 0.00024;
 Matches 107; Conservative 42; Mismatches 155; Indels 136; Gaps 21;
 265 HQEGHGSPEDINDREYKSRACASIQINEASDVSDSDVSDISSIDVSPDDVGI 324
 54 HRQGG-----ESGFAKRAISLVKLEKREDEL--DSITAITNGAHPKCVII 102
 325 LGGRFRWRARKATANQ---RVFAVLQF---ELHR-LIKVQLIAASPDILLDEISFLGK 377

103 ---QRTLDGRLOVAGRGKGFPHVYARIWRPDLHKNELKHVKYCAFAFDLKCDSVC--- 155
 378 VSAKSYFVKLLPSEFLYKPLP-HVVVKQRGDSKTDQHKMSSAENVVRLSNQG--- 433
 156 VNPYHY-----ERVSPGIDLSGLSLQSGPSRLV---KDEYSAGPLVGSMDIDGNDI 204
 434 ---HHQSQSNYMPFANNPPASPAENGYCFPPQPPPSGNHQWLIPVSPSEGLIYKPH-- 487
 205 GTIQHH-----PTQMVGPGGYGYP-----QGPSEYAV--PHGM 235
 488 PGVAHTGHYGYHYMPTPMVMPQVHPGMPGPPGNGYFPYVGMPTIMNPNVCSQQQQ 547
 236 PGMP-----GPMNPGVMAPPPPPQQAQNPQNGVHHTQANSFTDPASALAMQQQQ 286
 548 QQPNEOMNFGHPGNLQNTQQQQRSD-----NEPAPQOQQQ-- 585
 287 QQQQQQQQQQ-----QQQQQSGVPGNSVAGGAAGGQYGGPPVPSQMQGA 336
 586 ---PTKSYPRARK-----SRQSTGSPSPGPGISGSKSFRFAAVIDESNINNAPEQT 636
 337 GGGGTSVAFSVHAQQNGYVQPGSAGSAPVGGGSGVFTAQPTP-----QQPQQP 385
 637 MTTTTTTRTVTQTRDGG 656
 386 PTGVQANTGSAGAGAGGG 405
 RESULT 20
 Q9V6W7 PRELIMINARY; PRT; 1594 AA.
 ID Q9V6W7 Q9V6W8;
 AC Q9V6W7; Q9V6W8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG8118 protein.
 GN MAM OR CG8118.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champagne M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jaiswal M., Kalish J., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleub J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195 (2000).
RA [2]
RP SEQUENCE FROM N.A.
RA Celnikier S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Parag V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RA [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kanink J.S., Prochink S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman C., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RA [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RA [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL: AE003816; AAF58300.2; -.
RA FlyBase; FBgn002643; mam.
RA GO: GO:0005634; C:nucleus; IDA.
RA GO: GO:0007500; P:mesoderm cell fate determination; IMP.
RA GO: GO:0007399; P:neurogenesis; NAS.
SQ SEQUENCE 1594 AA; 167262 MW; D4236DA26F70D092 CRC64;

Query Match
Best Local Similarity 4.9%; Score 180; DB 5; Length 1594;
Matches 162; Conservative 95; Mismatches 306; Indels 328; Gaps 35;

QY 40 EQLSPSQRFQDHGTMTNSRNNSTLTIVHPGSPQPCGVNLSVOLDSSAQAQTEKFV 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 QETTVLQKFLFLE--SKNKAARKTKDKLPDPQHQHQHQHQHQHQHQHQHQHQ 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 100 SQMSFMENVRSSAQHDQRKVRREED-----FAVPVYINRRSRSGHRTKSGIEKEKH 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 GQL-----QSSVHVQKQFLKPAEDVDNGPDSPFPHKLNNNNNSNNNNNANAN 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 153 TPMPAPSSHSHTRFQVNTQSGKQNVCLATCSKPEVRDQVKANARSGGFVILDSVTBE 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 -----GGNGSNTGNTNNNGNSTN-----NNGGSGNNNGSENLTKFSVEIVQ 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 213 IDEKSA--SSHDVRND-----YNASLRQE-----SRNLYRDGGKT 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 LEFTTSAANSQQQISTNTVTVKALNTNTSVKSPGVGGGGGGGGGNNNNNGGGGGG 379

```


01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
Hypothetical protein KIAA1816 (Fragment).
KIAA1816.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISUB=Brain.
MEDLINE=21245130; PubMed=11347906;
Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
"Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
DNA Res. 8:85-95(2001).
EMBL; AB058719; BAB47445.1; --
Genew; HGNC:16272; MAMU3.
Hypothetical protein.
NON TER 1
SEQUENCE 1162 AA; 124759 MW; 0447AF72DC0B877E CRC64;
Query Match 4.9%; Score 179.5; DB 4; Length 1162;
Best Local Similarity 19.0%; Pred No. 00073;
Matches 135; Conservative 91; Mismatches 232; Indels 253; Gaps 30;
QY 125 DPAVP-----VYINRRSQSGHRTKSGIEKXHTMPWAPSSH-----162
DB 32 DFAAPAAAANGSSICINSLSSLGAGIGVNNTPNTPAAPSNHPAAGCGSGGPG 91
QY 163 -SIRFQEVNQTGSKQNVCLATCSKEVRDQVKANARSGFVLSLVTEEDLEKSASS 221
DB 92 GSAAPVKHSTVVERLQRTEGCRHHVNCENR-----YQQAQVEQLELER--- 136
QY 222 HFRVNDYNASLRQESRNLRYRDGKTRKLDTDNGAESHLATE-NHS-----QEG 269
DB 137 RDTVSLYQRTLEQRAKSGAGTGKQHPKSPQODAEASAEQNRHTLIMQETVKRLEG 196
QY 270 HGSPEDINDREYSKSRAC-ASLQIQINEASDDVSDMSVDSISSI-----DVSPPDV 321
DB 197 ARSP--LNGDQ--NGACDGNFSPSKIRKDIS--AGMEAINLNSMPLPSASP--- 246
QY 322 VGILGQKRFWRKAKAIAQORFAVOLFEHLR-----LKVKLIAASPDLLDEISFLG 376
DB 247 -----LHQDLKPSLPLQNSGTHTPG--LLEDLSKNG 276
QY 377 KVSASVYVK-----KLLPSEFLVKPPLPHVVVKRGDSEKTDQHKWESS----- 421
DB 277 RLPEIKLVNCGSDLEDSFTILQSKDLQEPDLPDPTCIDTSETSLNQKLFSDINLDQ 336
QY 422 -----AENVVGLSNQGHQOS-----NYMPEA 444
DB 337 EQQLDELANTVPEDDIQDLFNEDEFEKKEPFSQFATPTLSQESASVKSDFSPSPFA 396
QY 445 N-----NPPASPAFGYCFPP-----QPPSGNHQQLIPVMSPSGLIYKP 486
DB 397 HVSMGSPQARPSGSG---PPFSTVSTATSLPSVASTPAAPN-----PASSPANCAVQSP 447
QY 487 H-PCMAHTGHVGYGYHYMPTVMVMPQYHFGMGPFPPEGNGYF-----527
DB 448 QTPNQAIT-----FGAPPRFGNGYLLNPAAVTVAGSASGEVAV 486
QY 528 PPGMPTIMNYPYCSQQQQQQQNEQNMQFHPGNLQNTQQOQORSNDEPAQOQQOQPT 587
DB 487 PSSDMSPAEQLKQMAQQOQRAKLMOQKQ-----QQQQQQQQ-----QQQQQQQ 531
QY 588 KSYFRKRSQSGTSGSSPSGPG--ISGSKSFRFAADEVSDSNINN--APEQTMTTTTTT 644
DB 532 QQQQHSNOTSNWSPLGPPSPSYGAFTAEKPNSPMYPQAFNNQNPVPPMANNLQKTTM 591
QY 645 RTTVTQTTRDGGVTRIVKVVPHNAKLASENAARIFOSIQEERKRYDSSKP 695
DB 592 NNLYLPQNHN-----MINQCPNNLGTNSLNK-----QHNILTYGNTKP 629

DT
GN
DB
OS
OC
OX
RN
RP
RC
RX
RT
RT
RT
RT
RL
DR
DR
KW
FT
SQ

GO; GO:0004527; F:exonuclease activity; IEA.
InterPro; IPR004859; Put 53exo.
Pfam; PF03159; XRN_N; 1.
Exonuclease.
SEQUENCE 1749 AA; 200750 MW; 991F704A3592B2D6 CRC64;
Query Match 4.9%; Score 180; DB 5; Length 1749;
Best Local Similarity 18.1%; Pred. No. 0.0012;
Matches 141; Conservative 117; Mismatches 245; Indels 278; Gaps 36;
QY 6 DEEKLPEMFLHVNADKQ--GPRAPRN-----KMAIYEQLSIPS 46
DB 1111 ERHHLIKGLQKODLNNSSQNGKSTLSQNTSYFQSPANTGKFLHGDVRVSLDKGNLP- 1169
QY 47 QRFQDHGTWNSRNTSTLV-----HPGSSQP-COVERNLVSQHL-----DSS 89
DB 1170 ---FGTGTVASIQDQKVVDVTEFCFAGNSLDGYCSEKRGICISKLRLYNLSCPSPPKS 1227
QY 90 AANQATEKPFVQMSFMENVR-----SSAQHDQRKVRVEDDFAPVYVINSRQSGHRTKS 145
DB 1228 TINKFYDQSIDPAEYWEKVTQOQNNHNGQKKI-----YSGGQKL 1269
QY 146 GIEKEKHTMPWAPSSHHSIRFQEVNQTGSKQNVCLATCSKEVRDQVKANARSGFVLSL 205
DB 1270 NQOIQDQTPITNAVENKELNWQLO-----LNNISNQ-----1303
QY 206 DVSVTEEDLEKSASSHSDRVNDYNASLRQESRNLRYRDGKTRKLDTDNGAESHLATENH 265
DB 1304 -----QHANNNNNNNNYN-----NNNNNHGQNH 1329
QY 266 SQ-EGHSGPEDINDREYS--KSRACASLQINEEASDDVSDMSVDSISSIDVSPDDV 321
DB 1330 NQNKVNOHPLAINPNPSVNYPMKRTKYKQNEFQOEYADLS-----KNYPNLEYN--- 1380
QY 322 VGILGQKRFWRKAKAIAQORFAVOLFEHLRLIKVKLIAASPDLLDEISFLGVSAS 381
DB 1381 -----FYDQ-----NDQKQOQL-----QPKQOQOQPOQ-----POP 1412
QY 382 SYPVKLLPSEFLVKPPLPHVVVKRGDSEKTDQHKWESSAENVV-----GRLSN 431
DB 1413 KQPKQPKQ 1472
QY 432 QGHH-----QQSNMPEFA-----NNP-----PASAPN--GYCFPPQ 462
DB 1473 EHNGDGSSEQVQTNFALLNNMFASTSISDQNDPDGLDQGPFPNMGH-YPPGP 1531
QY 463 PPSGNHQWLIPVMSPSGLIYKHPGMAHTGHVGYGYHYMPTVMVMPQYHFG----- 516
DB 1532 PPM-----MGPPEH-----YHPGH-----SYPPPPPHMNGYPPGPPPHM 1567
QY 517 MGFPF---PGNGYPPYGMPTIMNYPYCSQQQQQQQNEQNMQFHPGNLQNTQ----- 568
DB 1568 MGYPHYHGHYPHPHPHGHQOEHHHQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1627
QY 569 -----QQQQ-----RSDNEPA-PQOQOQPTK-----SYPR-ARKSRQ 598
DB 1628 QPNQNTQPTQDQOQOQPPKQ 1687
QY 599 GSTGSSPGPGIGSK-----SFRFAADEVSDSNINNAPQMTTTTTTTTQTTTQTT 651
DB 1688 PAQPKQPAQKQ 1746
QY 652 T 652
DB 1747 T 1747

RESULT 22
Q96JK9 PRELIMINARY; PRT; 1162 AA.
AC Q96JK9
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```
RESULT 23
077069 ID 077069 PRELIMINARY; PRT; 697 AA.
AC 077069;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MEDA-B.
GN MED OR MEDEA OR CG1775.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98361878; PubMed=9694800;
RA Xu X., Yin Z., Hudson J.B., Ferguson E.L., Frasch M.;
RT "Smad proteins act in combination with synergistic and antagonistic
RT regulators to target Dpp responses to the Drosophila mesoderm.";
RL Genes Dev. 12:2354-2370(1998).
DR EMBL; AF13485; AAC35436.1; -.
DR HSP; Q13485; ID1.
DR TRANSFAC; T04382; -.
DR FlyBase; FBgn0011655; Med.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IGI.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; NAS.
DR GO; GO:0008283; P:cell proliferation; IMP.
DR GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR GO; GO:0007179; P:IGF1 receptor signaling pathway; IGI.
DR GO; GO:0007476; P:wing morphogenesis; IMP.
DR InterPro; IPR001132; Dwarfin.
DR InterPro; IPR003619; Dwarfin A.
DR InterPro; IPR008984; SMAD_FHA.
DR Pfam; PF03165; MH1; 1.
DR Pfam; PF03166; MH2; 1.
DR SMART; SM00523; DW4; 1.
DR SMART; SM00524; DWB; 1.
SQ SEQUENCE 697 AA; 73668 MW; 9C33C49E1B3ACC5C CRC64;

Query Match
Best Local Similarity 23.1%; Score 178.5; DB 5; Length 697;
Matches 116; Conservative 50; Mismatches 177; Indels 159; Gaps 23;

QY 265 HSEGHGSPEDINDREYKSRACASLQINEASDDVSDSDSISIDVSPDDVVG 324
Db 54 HQGG-----ESEGFAKRAIESLVKLEKRDDEL--DSLITAITNGAHPSCVTI 102

QY 325 LGOKREWRARKATANOQ---RVFAVOLF---ELHR-LIKVOKLIASPDLLIDEIFLKG 377
Db 103 ---QRTLDGLQVAGKRGPHVIYARIWPDLDHKNELKHVKYCAFAFDLKCDVSC--- 155

QY 378 VSAKSYVVKLLPSEFLVKKPLP-HVVVKQRGDSEKTDQHKMESSAENVVGRLSNQ--- 433
Db 156 VNPYH-----ERVSPGIDLSGLSLQSGFSRLV---KDEYSAGPLVGSMDIDGNDI 204

QY 434 ----HQQSNYMPFANNPPASPNAGYCFPPQPPSGNHOQWLIPIVMSPEGLYKPH-- 487
Db 205 GTIQHH-----PTQWVPGGYYGP-----QGFSEYAV--PHGM 235

QY 488 PGMAHTGHYGGYGHVYPMYPMQVHPGMGFPFPPGNGYPPYGMPTTINPCSSQOQO 547
Db 236 PGMP-----GPMNPGVWAPPPPPQOQNPQNGVHHHTQANSPTDPASALAMQOQO 286

QY 548 QQQPNEQMNQFGHPGNLQNTQQOQQRSD-----NEPAPQQOQO-- 585
Db 287 QQQOQQOQQO-----QQOQQOQSGVNGSVNAGGAAAGGYQYQPPPPVSMQGA 336
```

```
QY 586 ---ETKSYPRARK-----SRQSTGSSPSGPGQISGSKSFRPFAADEVDSNINNAPROT 636
Db 337 GGGGTSVAPSVHAQONGYVSQFGSAGAPVGGGVFGTA-----OFTSQPOOP 385

QY 637 MTTTTTTRITVTQTTRDGGG-----VTRVIKVVPHNAK-----LAS 673
Db 386 PTGVOANTGSAGAQAGAGGAAGTGTGNTLTITQSMQPPNPRSLPGGFWNSSLGDLGS 445

QY 674 ENAARIFQSIOERKRYDSSKP 695
Db 446 PQOTPPQOQQOQQPRLSRQP 467

RESULT 24
Q93YP4 ID Q93YP4 PRELIMINARY; PRT; 1024 AA.
AC Q93YP4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Epsin-like protein (At3g59290).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Jones T., Bath J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Hsuan W.W., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong H.C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV059878; AAL24360.1; -.
DR EMBL; BT008432; AAP37791.1; -.
DR InterPro; IPR001026; ENTH.
DR InterPro; IPR008942; ENTH_VHS.
DR Pfam; PF01417; ENTH; 1.
DR SMART; SM00273; ENTH; 1.
SQ SEQUENCE 1024 AA; 109694 MW; 1C8A5F33EB12ED04 CRC64;

Query Match
Best Local Similarity 19.2%; Score 178.5; DB 10; Length 1024;
Matches 151; Conservative 77; Mismatches 280; Indels 277; Gaps 33;

QY 84 QHLDSSAANOATEKFSQMSFMENVRSSAQHQDKMKVREEDFAVPIVYNNRRSQSHGRT 143
Db 116 QYIDSSGQOQS-----NVRKKAQ-----SIVALVNDKERITEVRE 151

QY 144 KSGIEKEK-HTPMVAPSSHHHSIRFQEVNQTGSKQNVCLATCSKPE--VRDQVKANARSG 200
Db 152 KAAANRDKYNSMHRPSGGYGDYEGRYGRDRDEGRSSYGKEREYGYRDDDR-NSRDGD 210

QY 201 FVISLDSVTEIDLEKSASSHDR-----VNDYNASL-RQESNRRLXR 242
Db 211 -----RYSRDSERYGRDNTDDEYGRSRSVDNYNGRSGSSDRERPLE 255

QY 243 DGGKTRLDKDTNGASHLATENHSGHSGSPEDINDREYKSRACASLQINEASDDV 302
Db 256 DDGQS--SSRDGSAFA-----DDHSQDGRGGLF-----RKFSQNTGAAPPSEAVSES 304
```

Thu Jul 29 08:52:25 2004

303 SD-DSMVDISSIDVSPDVGILGOKRFRARKA--TANQORVFAVOLFEHLRLIKYQK 359
 305 SPVYSERDGGEPQVAPPAGAAASPLAENISVDNKAADFVNESSPOQVEAF----- 354
 360 LIAASPDLLLDELISFLGKYSASYP-----VKLLPSEFLVKP----- 397
 365 -----DBFDPRGSVSAACAPTAGASVZAPIPPTVVSTPAPPASINAEMDLLGSL 404
 398 ----PLPHVVVKQRGDSEKTDQH-----KMESSAE-----NVVGLSNQ 432
 405 DVFSNPPLAIVSDSTSIVETNGQANTGLAPSFSTQSSTQPDPPDGGDFPFAITSADE 464
 433 GHHQOQNYMPFANNPPAS-----PANGYCFP----- 459
 465 TSHQSGFGVFPQPTPTSPNPNNEHFGGEAFSAVTDSEPGVQNMQAPPNLSVFPQEQD 524
 460 -----POPP-----PSGN-----HQ-----QWLI 473
 525 TSQSEIDILAGILPPSGPPVLSLPQDSTMTPTSQFHPNGNSYESYHHQAAPTDLNMQGT 584
 474 VMSPSGELIYKHPGMAHTG---HYGGY---GHYMPMPVMQVHPGMPGPPNGYF 527
 585 PFGQASQGFNMVSHSQNHHEGMQFNNGFTQPCYAGPATSQPPQYTPGVSSHPPSES-F 643
 528 P-----PYGMPTI-----MNPYCSSQOQO-----QQQNEQM 555
 644 PHPGSATSASSQTPYATTPNVSAQFDGSGFMTQQPYGVTOQVHVVPVSHIPORTSGPV 703
 556 NQCHPNLQNTQOQ-----QORSNEPAPQOQOQTKSYPRARKSRQSGTSSP---SGP 608
 704 AAFGNNNIVGDHMQPSTSSSQTPYPTPNAPSQFDGNGFMTQPYGVIPQVHGVP 763
 609 QGISGSKSRFPFAVDEDSINN---APEQMTTTTTTTTQTTTQTTDGGGVTRVKKV 665
 764 SHIPQRTSGFVAAGHNSNVVGMFSPAGLSLETSSASQPSLTP-----LTGATEIV 816
 666 PHNAK 670
 817 PQNOK 821

RESULT 25
 001505 PRELIMINARY; PRT: 1239 AA.
 AC 001505;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN C37A2.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RL investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Le T.T., Kemp K., Scheet P.;
 RT "The sequence of C. elegans cosmid C37A2.";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RT "Direct Submission.";

Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U97194; AAB52447.2; -.
 DR PIR: T30160; T30160.
 DR WormPep: C37A2.2; CE27747.
 KW Hypothetical protein.
 SQ SEQUENCE 1239 AA; 135436 MW; 81F6C299FDC77D85 CRC64;
 Query Match 4.9%; Score 178.5; DB 5; Length 1239;
 Best Local Similarity 23.4%; Pred. No. 0.00095;
 Matches 137; Conservative 66; Mismatches 235; Indels 147; Gaps 32;
 QY 150 EKHTPWAPSSHHSTRFQ-EVNOTGSKQNV-CLATCS-----KPEVRDQVKANARSGGFV 202
 Db 20 EKVGDIVRPSATALLSSLTSSSGIEASTCASTPTTANTLPLLEEESVEDEYADGVG 79
 QY 203 ISLDVSVTEIDLEKSASSHDRV-----NDYNASLRFQESNRNRLYRDGKTRLKDTDNGAE 257
 Db 80 IATE-KLVKSLIVFVPSDTARVTSAINDTVASGKTNEEDR-----KSIDTSVD 128
 QY 258 SHLATENHSQEGHSGPEDIDN---DREYSKSRACASLQ--QINERASDDVSDSMVDSIS 312
 Db 129 --LIQEREEQEDTG-----DNQISQSSSKDHECPEREDLSIHEEPTNFTNDGVLPSTS 181
 QY 313 S-IDVSPDDVVGI-----LGQKRFWRARKAIAQ--ORVFAVOLFEHLRL-----IKVQ--- 358
 Db 182 SGIPASPADMTASSRNPLRKRVYSLRLPAKAAQLRIANFQGSRRIRIGLSALKISDMD 241
 QY 359 --KLIAASPD-----LLLDLDEISFLGKVSASYPVKKL---LPSBFLVKPPLP-----HVVK 405
 Db 242 CLRADGSPPOI PVALEDHINISTQENPRDIRLSLHQLPMTIPQOQPPPPQVPSHMLPS 301
 QY 406 QRGSEKTDQHKVMESSAENVVRLSNQGH-----QSNYMPFANNP-PASPAINGYCF 458
 Db 302 STQOPGHPSHH-MGPLSQQLLPSGPNQGHHSFQVMVKQEPSPQTPQPHMQOTPOQOVL 360
 QY 459 PPQPPPSGNHQOQWLI PVMSPSEGLIYKPHPG-MAHTGHYGGYGYHYMPTMVMMPQYHPGM 517
 Db 361 PQYPPGMPQPHMQHMQMTAEVYAQMRARSGFMAAQIKQEVPSGSGQTPV-----PGT 414
 QY 518 GFP---PPGNGYFPYGMW---PTI-----MNPYCSSQOQOQ----- 548
 Db 415 PQOQIITPQFSLGPMGSLGPPTAPPQSQPMNPQOQRIQOQOQAAPSASNSPLLVNLSN 474
 QY 549 QCPNEQMNQFCHPGN-----LQNTQOQOQSDNEPAPQ 582
 Db 475 QQPPOQ--QYMPGFSAGLSMQIAAIQOQOQHQYQQRILQOQOQOQAMQOQOQOQVQ 532
 QY 583 QOQ-----PTKSYPRAKSRQSGTSGSPSGPQGISGSKSRPEA 621
 Db 533 QOQAPPTTTPNPGHP-----QGFPTNQAGPGTPTG-RPIPPYA 570

Search completed: July 29, 2004, 08:24:47
 Job time : 88 secs